A Bayesian-motivated test for high-dimensional linear model with fixed design

Rui Wang[1] and Xingzhong Xu[*1,2]

[1] School of Mathematics and Statistics, Beijing Institute of Technology, Beijing 100081,China

[2] Beijing Key Laboratory on MCAACI, Beijing Institute of Technology, Beijing 100081,China

Sunday 10th March, 2019

Abstract

This paper considers testing regression coefficients in high-dimensional linear model with fixed design matrix. We prove that there does not exist nontrivial unbiased test for this problem. This phenomenon makes it impossible to consider the problem from a minimax perspective. Nevertheless, Bayesian methods can still produce tests with good average power behavior. We propose a new test statistic which is the limit of Bayes factors under normal distribution. The null distribution of the proposed test statistic is approximated by Lindeberg's replacement trick. Under certain conditions, the global asymptotic power function of the proposed test is also given. The finite sample performance of the proposed test is demonstrated via simulation studies.

Key words: High-dimensional test, Lindeberg method, Linear model, Unbiasedness.

1 Introduction

Consider linear regression model of the form

$$\mathbf{y} = \mathbf{X}_a\boldsymbol{\beta}_a + \mathbf{X}_b\boldsymbol{\beta}_b + \boldsymbol{\epsilon}, \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^n$ is the response, $\mathbf{X}_a$, $\mathbf{X}_b$ are $n \times q$ and $n \times p$ design matrices, respectively, $\boldsymbol{\beta}_a \in \mathbb{R}^q$, $\boldsymbol{\beta}_b \in \mathbb{R}^p$ are unknown regression coefficients, and $\boldsymbol{\epsilon} \in \mathbb{R}^n$ is the vector of random errors. Here the

*Corresponding author

Email address: xuxz@bit.edu.cn

design matrix $\mathbf{X}_a$ contains the predictors that are known to have effects on the response, and our interest is to test if $\mathbf{X}_b$ contains any useful predictors. That is, we would like to test the hypotheses

$$\mathcal{H}_0 : \boldsymbol{\beta}_b = 0, \quad \text{v.s.} \quad \mathcal{H}_1 : \boldsymbol{\beta}_b \neq 0. \tag{2}$$

The conventional test for hypotheses (2) is the F-test which is also the likelihood ratio test under normal errors. However, the F-test is not well defined in high dimensional setting. In fact, if $\boldsymbol{\epsilon}$ is normal distributed and $\text{Rank}[\mathbf{X}_a; \mathbf{X}_b] = n$, then the likelihood is unbounded under the alternative hypothesis. Thus, new test methodology is required in high-dimensional setting.

For the problem of testing hypotheses (2), two different high-dimensional settings have been extensively considered in the literature. One is the small p, large q setting. An important example of this setting is testing individual coefficients of a high-dimensional regression. See (Bühlmann, 2013; Zhang and Zhang, 2014; Lan et al., 2016b) for testing procedures in this setting. In this paper, however, we focus on the other setting, namely the large p, small q setting. In this case, there are just a few covariates, namely $\mathbf{X}_a$, are known to have effects on the response, while there remain a large number of covariates, namely $\mathbf{X}_b$, to be tested.

Many test procedures have been proposed in the large p, small q setting. Based on an empirical Bayes model, Goeman et al. Goeman et al. (2006) and Goeman et al. Goeman et al. (2011) proposed a score test statistic as well as a method to determine the critical value. This score test was further investigated by Lan et al. Lan et al. (2014) and Lan et al. Lan et al. (2016a). Based on U-statistics, Zhong and Chen Zhong and Chen (2011) proposed a test for the case where $\mathbf{X}_a = \mathbf{1}_n$. Later, Wang and Cui Wang and Cui (2015) extended the test of Zhong and Chen Zhong and Chen (2011) to the case where $\mathbf{X}_a$ is a general design matrix. To accommodate outlying observations and heavy-tailed distributions, Feng et al. Feng et al. (2013) proposed a rank-based test for the entire coefficients. Xu Xu (2016) modified the test of Feng et al. Feng et al. (2013) and proposed a scalar invariant rank-based test. Janson et al. Janson et al. (2016) utilized a convex optimization program to obtain an unbiased estimator of a signal to noise ratio with small variance, based on which a test is constructed. Apart from the afore mentioned tests, there is another line of research utilizing desparsified Lasso estimator; see Dezeure et al. (2017) and the references therein.

Except for the test of Goeman et al. Goeman et al. (2006), most existing high dimensional tests adopted the random design assumption, that is, the rows of $\mathbf{X}_b$ are considered as being generated from a super population. As noted by Lei et al. Lei et al. (2018), assuming a fixed design or a random design could lead to qualitatively different inferential results and the former is preferable from a theoretical point of view. Hence we focus on the fixed design setting in this paper. Of course, our results are still valid for random design from a conditional inference perspective.

In Bayesian literature, many Bayesian tests have been proposed for hypothesis (2) in low-dimensional setting; see Javier Girón et al. (2006); Goddard and Johnson (2016); Zhou and Guan (2018) and the references therein. So far, most existing Bayesian tests are not applicable in large p, small q setting. In principle, using Bayes factor as a frequentist test statistic in high-dimensional setting is a good strategy for at least two reasons. First, the Bayes factors corresponding to proper priors are always well defined, even if the likelihood is unbounded. Second, under mild conditions, tests based on Bayes factors are admissible (See, e.g., Lehmann and Romano (2005), Theorem 6.7.2). In fact, the test procedure of Goeman et al. Goeman et al. (2006) is motivated by Bayesian methods but is treated as a frequentist significance test.

In this paper, we propose a new test statistic in large p, small q setting which is the limit of Bayes factors under normal linear model. The proposed test statistic is a ratio of quadratic forms of $\mathbf{y}$. We give an approximation of the distribution of quadratic form using Lindeberg's replacement trick. Based on this approximated distribution, the critical value of the proposed test statisitc is calculated by a one step iteration. We prove that the proposed test procedure is valid under weak conditions. In particular, we do not require that the test statistic is asymptotically normally distributed. Under certain conditions, we also derive the asymptotic power function of the proposed test and the test of Goeman et al. Goeman et al. (2006). It is shown that the proposed test can detect the signals from more directions than the test of Goeman et al. Goeman et al. (2006). Simulations are conducted to examine the performance of the proposed test.

The rest of the paper is organized as follows. In Section 2, we propose a Bayesian-motivated test and study the theoretical properties of the proposed test procedure. The asymptotic power function is also derived. Section 3 contains the simulation results. Section 4 concludes the paper. The technical proofs are presented in Appendix.

2 Test procedure

2.1 A Bayesian-motivated test

In this paper, we consider testing hypotheses (2) in large p, small q setting. More specifically, we make the following assumption throughout the paper.

Assumption 1. *In the linear model* (1)*, suppose* $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^\top$ *where* $\epsilon_1, \ldots, \epsilon_n$ *are independent indentically distributed (iid) with mean* 0 *and covariance* $\sigma^2 = \phi^{-1}$*. Furthermore, suppose* $q < n$*,* $\operatorname{Rank}(\mathbf{X}_a) = q$ *and* $\operatorname{Rank}([\mathbf{X}_a; \mathbf{X}_b]) = n$*.*

Testing hypotheses (2) in large p, small q setting is a challenging problem. Goeman et al. Goeman et al. (2006) noticed that their test has negligible power for many alternatives and consequently is not an unbiased test. Biased tests are often regarded as problematic in classical statistics. However, the situation is very different for our problem. In fact, we shall show that under normal assumption, there is no nontrivial unbiased test in large p, small q setting.

Let $\mathbf{P}_a = \mathbf{X}_a(\mathbf{X}_a^\top \mathbf{X}_a)^{-1}\mathbf{X}_a^\top$ be the projection matrix onto the column space of $\mathbf{X}_a$ and denote $\tilde{\mathbf{P}}_a = \mathbf{I}_n - \mathbf{P}_a$. Define $\boldsymbol{\theta} = (\boldsymbol{\beta}_a^\top, \boldsymbol{\beta}_b^\top, \phi)^\top$ and

$$\Theta_0 = \left\{\theta = (\boldsymbol{\beta}_a^\top, \boldsymbol{\beta}_b^\top, \phi)^\top : \boldsymbol{\beta}_a \in \mathbb{R}^q, \boldsymbol{\beta}_b = 0, \phi > 0\right\},$$
$$\Theta = \left\{\theta = (\boldsymbol{\beta}_a^\top, \boldsymbol{\beta}_b^\top, \phi)^\top : \boldsymbol{\beta}_a \in \mathbb{R}^q, \boldsymbol{\beta}_b \in \mathbb{R}^p, \phi > 0\right\}.$$

A test function $\varphi(\mathbf{y})$ of level α is a Borel measurable function from $\mathbb{R}^n$ to $\mathbb{R}$ such that $0 \le \varphi(\mathbf{y}) \le 1$ and $\mathrm{E}_\theta[\varphi(\mathbf{y})] \le \alpha$ for $\theta \in \Theta_0$. A test function $\varphi(\mathbf{y})$ is nonrandom if $\varphi(\mathbf{y}) \in \{0, 1\}$. Let $\lambda(\cdot)$ be the Lebesgue measure on $\mathbb{R}^n$. We have the following proposition.

Proposition 1. *Suppose Assumption 1 holds. Furthermore, assume* $\boldsymbol{\epsilon} \sim \mathcal{N}_n(0, \phi^{-1}\mathbf{I}_n)$. *Let* $\varphi(\mathbf{y})$ *be a test function of level* α, $0 < \alpha < 1$. *Then the following results hold.*

(a) If $\varphi(\mathbf{y})$ *is unbiased, that is,* $\mathrm{E}_\theta[\varphi(\mathbf{y})] \ge \alpha$ *for* $\theta \in \Theta$, *then* $\varphi(\mathbf{y}) = \alpha$, *a.s.* λ.

(b) If $\varphi(\mathbf{y})$ *is nonrandom, then for every* $M > 0$, *there exists a parameter* $\theta = (\boldsymbol{\beta}_a^\top, \boldsymbol{\beta}_b^\top, \phi)^\top \in \Theta$ *such that* $\phi \boldsymbol{\beta}_b^\top \mathbf{X}_b^\top \tilde{\mathbf{P}}_a \mathbf{X}_b \boldsymbol{\beta}_b > M$ *and* $\mathrm{E}_\theta[\varphi(\mathbf{y})] < \alpha$.

In (b) of Proposition 1, the quantity $\phi \boldsymbol{\beta}_b^\top \mathbf{X}_b^\top \tilde{\mathbf{P}}_a \mathbf{X}_b \boldsymbol{\beta}_b$ can be regarded as a signal to noise ratio. In any conventional sense, it is expected that the power of a reasonable test should tend to 1 as $\phi \boldsymbol{\beta}_b^\top \mathbf{X}_b^\top \tilde{\mathbf{P}}_a \mathbf{X}_b \boldsymbol{\beta}_b \to \infty$. Indeed, such a test can be easily constructed in low dimensional setting. However, (b) of Proposition 1 claims that it is impossible to find such a test in large p, small q setting. Thus, it is impossible to consider the problem from a minimax perspective. This motivates us to adopt Bayesian methods to find a test with good average power behavior.

Bayes factor (see, e.g., (Kass and Raftery, 1995)) is a commonly used tool for model comparison within the Bayesian framework. In our problem, suppose $\boldsymbol{\epsilon} \sim \mathcal{N}_n(0, \phi^{-1}\mathbf{I}_n)$, the Bayes factor for hypotheses (2) is

$$B_{10} = \frac{\int d\mathcal{N}_n(\mathbf{X}_a\boldsymbol{\beta}_a + \mathbf{X}_b\boldsymbol{\beta}_b, \phi^{-1}\mathbf{I}_n)(\mathbf{y})\pi_1(\boldsymbol{\beta}_b, \boldsymbol{\beta}_a, \phi)\,\mathrm{d}\boldsymbol{\beta}_b\mathrm{d}\boldsymbol{\beta}_a\mathrm{d}\phi}{\int d\mathcal{N}_n(\mathbf{X}_a\boldsymbol{\beta}_a, \phi^{-1}\mathbf{I}_n)(\mathbf{y})\pi_0(\boldsymbol{\beta}_a, \phi)\,\mathrm{d}\boldsymbol{\beta}_a\mathrm{d}\phi},$$

where $d\mathcal{N}_n(\mu, \boldsymbol{\Sigma})(\mathbf{y})$ is the density function of a $\mathcal{N}_n(\mu, \boldsymbol{\Sigma})$ random vector evaluated at $\mathbf{y}$, $\pi_0(\boldsymbol{\beta}_a, \phi)$ and $\pi_1(\boldsymbol{\beta}_b, \boldsymbol{\beta}_a, \phi)$ are the prior densities under the null and alternative hypotheses, respectively. If B_{10} is large, the alternative hypothesis is preferred. It is known that if the true parameters indeed come from the specified prior distribution, then the Bayes factor is the likelihood ratio statistic and consequently, the test based on the Bayes factor is the most powerful test. We notice that in the frequentist framework, the Bayes factor can be treated as a test statisitc, provided the critical value can be determined to preserve the test level.

The behavior of a Bayes factor largely depends on the choice of priors. In Bayesian literature, many priors have been considered for testing the coefficients of linear model. Popular priors include g-priors ((Liang et al., 2008)) and intrinsic priors ((Casella and Moreno, 2006)). Unfortunately, these priors are not well defined in large p, small q setting. Note that under the null hypothesis $\mathcal{H}_0$, the model is low-dimensional. This allows us to impose the reference prior $\pi_0(\boldsymbol{\beta}_a, \phi) = c/\phi$, where c is a constant. Under $\mathcal{H}_1$, write $\pi_1(\boldsymbol{\beta}_b, \boldsymbol{\beta}_a, \phi) = \pi_1(\boldsymbol{\beta}_b|\boldsymbol{\beta}_a, \phi)\pi_1(\boldsymbol{\beta}_a, \phi)$. For parameters $\boldsymbol{\beta}_a$ and ϕ, we consider the same prior as in $\mathcal{H}_0$, that is $\pi_1(\boldsymbol{\beta}_a, \phi) = \pi_0(\boldsymbol{\beta}_a, \phi)$. For parameter $\boldsymbol{\beta}_b$, however, imposing the improper reference prior would not produce valid marginal density of $\mathbf{y}$. To make the marginal density of $\mathbf{y}$ well defined, we consider the simple normal prior $p_1(\boldsymbol{\beta}_b|\boldsymbol{\beta}_a, \phi) =$

$d\mathcal{N}_p(0, \kappa^{-1}\phi^{-1}\mathbf{I}_p)(\boldsymbol{\beta}_b)$, where $\kappa > 0$ is a hyperparameter. That is, we put the following priors,

$$\pi_0(\boldsymbol{\beta}_a, \phi) = \frac{c}{\phi}, \quad \pi_1(\boldsymbol{\beta}_b, \boldsymbol{\beta}_a, \phi) = \frac{c}{\phi} d\mathcal{N}_p\left(0, \frac{1}{\kappa\phi}\mathbf{I}_p\right)(\boldsymbol{\beta}_b). \tag{3}$$

Let $B_{10,\kappa}$ be the Bayes factor corresponding to the priors (3). It is straightforward to show that

$$\begin{aligned} 2\log(B_{10,\kappa}) =& p\log\kappa - \log|\mathbf{X}_b^\top \tilde{\mathbf{P}}_a \mathbf{X}_b + \kappa\mathbf{I}_p| \\ & - (n-q)\log\left(1 - \frac{\mathbf{y}\top\tilde{\mathbf{P}}_a\mathbf{X}_b\left(\mathbf{X}_b^\top\tilde{\mathbf{P}}_a\mathbf{X}_b + \kappa\mathbf{I}_p\right)^{-1}\mathbf{X}_b^\top\tilde{\mathbf{P}}_a\mathbf{y}}{\mathbf{y}\top\tilde{\mathbf{P}}_a\mathbf{y}}\right). \end{aligned}$$

Denote by $\tilde{\mathbf{P}}_a = \tilde{\mathbf{U}}_a\tilde{\mathbf{U}}_a^\top$ the rank decomposition of $\tilde{\mathbf{P}}_a$, where $\tilde{\mathbf{U}}_a$ is a $n \times (n-q)$ column orthogonal matrix. Let $\mathbf{X}_b^* = \tilde{\mathbf{U}}_a^\top\mathbf{X}_b$, $\mathbf{y}^* = \tilde{\mathbf{U}}_a^\top\mathbf{y}$. Let γ_i be the ith largest eigenvalue of $\mathbf{X}_b^*\mathbf{X}_b^{*\top}$, $i = 1, \ldots, n-q$. Denote by $\mathbf{X}_b^* = \mathbf{U}_b^*\mathbf{D}_b^*\mathbf{V}_b^{*\top}$ the singular value decomposition of $\mathbf{X}_b^*$, where $\mathbf{U}_b^*$, $\mathbf{V}_b^*$ are $(n-q)\times(n-q)$ and $p \times (n-q)$ column orthogonal matrices, respectively, and $\mathbf{D}_b^* = \mathrm{diag}(\sqrt{\gamma_1}, \ldots, \sqrt{\gamma_{n-q}})$. Then we have

$$\begin{aligned} 2\log(B_{10,\kappa}) =& (n-q)\log\kappa - \sum_{i=1}^{n-q}\log(\gamma_i + \kappa) \\ & - (n-q)\log\left(1 - \frac{\mathbf{y}^{*\top}\mathbf{U}_b^*\mathbf{D}_b^*\left(\mathbf{D}_b^{*2} + \kappa\mathbf{I}_{n-q}\right)^{-1}\mathbf{D}_b^*\mathbf{U}_b^{*\top}\mathbf{y}^*}{\mathbf{y}^{*\top}\mathbf{y}^*}\right). \end{aligned}$$

The main part of the above expression is

$$T_\kappa = \frac{\mathbf{y}^{*\top}\mathbf{U}_b^*\mathbf{D}_b^*\left(\mathbf{D}_b^{*2} + \kappa\mathbf{I}_{n-q}\right)^{-1}\mathbf{D}_b^*\mathbf{U}_b^{*\top}\mathbf{y}^*}{\mathbf{y}^{*\top}\mathbf{y}^*}.$$

A large value of T_κ supports the alternative hypothesis. Hence T_κ can be regarded as a frequentist test statistic. It remains to choose an appropriate hyperparameter κ. As Goeman et al. Goeman et al. (2006) noted, the priors should place most probability on the alternatives which are perceived as more interesting to detect. Their strategy is to let the prior magnitude tend to zero to obtain a test with good power behavior under local alternatives, that is, $\|\boldsymbol{\beta}_b\|$ is small. Note that the prior magnitude of $\boldsymbol{\beta}_b$ decreases with the increase of κ. In fact, if we let κ tends to infinity, the limit

$$\lim_{\kappa\to\infty}\kappa T_\kappa = \frac{\mathbf{y}^{*\top}\mathbf{X}_b^*\mathbf{X}_b^{*\top}\mathbf{y}^*}{\mathbf{y}^{*\top}\mathbf{y}^*}$$

is exactly the test statistic of Goeman et al. Goeman et al. (2006).

As implied by Proposition 1, however, testing hypotheses (2) in large p, small q setting is a difficult problem and no unbiased test can guarantee large power, even if the signal is strong. Hence it may be too ambitious to consider local power behavior where signal is weak. Thus, contrary to the strategy of Goeman et al. Goeman et al. (2006), we let κ tend to 0 to obtain a test with good average power behavior for large $\|\boldsymbol{\beta}_b\|$. Note that the statistic T_κ degenerates to 1 as $\kappa \to 0$.

Nevertheless, the scaled statistic $(T_\kappa - 1)/\kappa$ has a proper limit as $\kappa \to 0$. Thus, we proposed the following test statistic

$$T = \lim_{\kappa \to 0} \frac{T_\kappa - 1}{\kappa} = -\frac{\mathbf{y}^{*\top}(\mathbf{X}_b^* \mathbf{X}_b^{*\top})^{-1}\mathbf{y}^*}{\mathbf{y}^{*\top}\mathbf{y}^*}.$$

The null hypothesis will be rejected if T is large.

2.2 Critical value

To formulate a valid frequentist test, we need to determine the critical value of T. If $\boldsymbol{\epsilon}$ were indeed normally distributed, then under the null hypothesis, $T \sim -(\sum_{i=1}^{n-q} \gamma_i^{-1} z_i^2)/(\sum_{i=1}^{n-q} z_i^2)$, where $z_1, \ldots, z_{n-q}$ are iid $\mathcal{N}(0,1)$ random variables. In this case, the exact critical value can be easily obtained. However, normal distribution rarely appears in practice. We would like to derive an asymptotic valid critical value for T_n for general distributions of $\boldsymbol{\epsilon}$.

Under the null hypothesis,

$$T = -\frac{(\sqrt{\phi}\boldsymbol{\epsilon})^\top \tilde{\mathbf{U}}_a (\mathbf{X}_b^* \mathbf{X}_b^{*\top})^{-1} \tilde{\mathbf{U}}_a^\top (\sqrt{\phi}\boldsymbol{\epsilon})}{(\sqrt{\phi}\boldsymbol{\epsilon})^\top \tilde{\mathbf{U}}_a \tilde{\mathbf{U}}_a^\top (\sqrt{\phi}\boldsymbol{\epsilon})}.$$

The numerator and the denominator of the above expression are both quadratic forms of iid random variables with mean 0 and variance 1. Hence the key step towards the goal is to approximate the distribution of the quadratic form of iid standardized random variables. The asymptotics of quadratic form have been extensively studied; see, e.g., Jiang (1996); Bentkus and Götze (1996); Götze and Tikhomirov (2002); Dicker and Erdogdu (2017); Bai et al. (2018). Most existing work use normal distribution to approximate the distribution of the quadratic form. However, normal distribution is just one of the possible limit distributions of quadratic form. See Sevast'yanov Sevast'yanov (1961) for a full characterization of the limit distributions of quadratic form of normal random variables. Our approximation strategy is to replace the random variables in quadratic form by suitable normal random variables. The error bound of this approximation will be derived by Lindeberg's replacement trick (see, e.g., Pollard (1984), Section III.4).

Let $\mathscr{C}^4(\mathbb{R})$ denote the class of all bounded real functions on $\mathbb{R}$ having bounded, continuous kth derivatives, $0 \le k \le 4$. It is known that if $\mathrm{E} f(Z_n) \to \mathrm{E} f(Z)$ for every $f \in \mathscr{C}^4(\mathbb{R})$ then $Z_n \rightsquigarrow Z$; see, e.g., Pollard (1984), Theorem 12 of Chapter III.

We have the following approximation theorem.

Theorem 1. *Let $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)^\top$, where ξ_i's are iid random variable with $\mathrm{E}\,\xi_1 = 0$, $\mathrm{Var}(\xi_1) = 1$. Furthermore, suppose the distribution of ξ_1 is symmetric about 0 and has finite eighth moments. Let $\mathbf{A}$ be an $n \times n$ symmetric matrix with elements $a_{i,j}$. Define*

$$S = \frac{\boldsymbol{\xi}^\top \mathbf{A} \boldsymbol{\xi} - \mathrm{tr}(\mathbf{A})}{\sqrt{2\,\mathrm{tr}(\mathbf{A}^2) + (\mathrm{E}(\xi_1^4) - 3)\,\mathrm{tr}(\mathbf{A}^{\circ 2})}},$$

where $\mathbf{A}^{\circ 2} = \mathbf{A} \circ \mathbf{A}$ and $\circ$ means Hadamard product. Let $z_1, \ldots, z_n$ be iid random variables with distribution $\mathcal{N}(0,1)$ and $\check{z}_1, \ldots, \check{z}_n$ be iid random variables with distribution $\mathcal{N}(0,1)$ which are in-

dependent of $\xi_1, \dots, \xi_n$. Let τ be a real number. Define

$$S_\tau^* = \frac{\tau \sum_{i=1}^n a_{i,i}\check{z}_i + 2\sum_{1\le i<j\le n} a_{i,j} z_i z_j}{\sqrt{2\operatorname{tr}(\mathbf{A}^2) + (\mathrm{E}(\xi_1^4) - 3)\operatorname{tr}(\mathbf{A}^{\circ 2})}}.$$

Then for every $f \in \mathscr{C}^4(\mathbb{R})$,

$$\begin{aligned}
&|\mathrm{E}\, f(S) - \mathrm{E}\, f(S_\tau^*)| \\
\le&\frac{\left|\mathrm{E}(\xi_1^4) - 1 - \tau^2\right|}{2}\|f^{(2)}\|_\infty \frac{\operatorname{tr}(\mathbf{A}^{\circ 2})}{2\operatorname{tr}(\mathbf{A}^2) + (\mathrm{E}(\xi_1^4) - 3)\operatorname{tr}(\mathbf{A}^{\circ 2})} \\
&+ \frac{\max\left(\left|\mathrm{E}(\xi_1^2 - 1)^3\right|, 12(\mathrm{E}(\xi_1^4) - 1)\right)}{6}\|f^{(3)}\|_\infty \frac{\sum_{l=1}^n \left(|a_{l,l}| \sum_{i=1}^n a_{i,l}^2\right)}{\left(2\operatorname{tr}(\mathbf{A}^2) + (\mathrm{E}(\xi_1^4) - 3)\operatorname{tr}(\mathbf{A}^{\circ 2})\right)^{3/2}} \\
&+ \frac{16\,\mathrm{E}(\xi_1^8) + 80\,\mathrm{E}(\xi_1^4) + 3\tau^4 + 96}{24}\|f^{(4)}\|_\infty \frac{\sum_{l=1}^n \left(\sum_{i=1}^n a_{i,l}^2\right)^2}{\left(2\operatorname{tr}(\mathbf{A}^2) + (\mathrm{E}(\xi_1^4) - 3)\operatorname{tr}(\mathbf{A}^{\circ 2})\right)^2}.
\end{aligned} \tag{4}$$

Remark 1. If $\tau^2 = \mathrm{E}(\xi_1^4) - 1$, the first term of the right hand side of (4) disappear. In practice, however, the quantity $\mathrm{E}(\xi_1^4)$ is often unknown and τ^2 should be chosen as an estimator of $\mathrm{E}(\xi_1^4) - 1$.

Remark 2. As noted in Chatterjee Chatterjee (2008), Section 3.1, an almost necessary condition for the asymptotic normality of S is

$$\frac{\operatorname{tr}(\mathbf{A}^4)}{\left(2\operatorname{tr}(\mathbf{A}^2) + (\mathrm{E}(\xi_1^4) - 3)\operatorname{tr}(\mathbf{A}^{\circ 2})\right)^2} \to 0. \tag{5}$$

On the other hand, it can be seen that the right hand side of (4) converges to 0 provided $\tau^2 = \mathrm{E}(\xi_1^4) - 1$ and

$$\frac{\sum_{l=1}^n \left(\sum_{i=1}^n a_{i,l}^2\right)^2}{\left(2\operatorname{tr}(\mathbf{A}^2) + (\mathrm{E}(\xi_1^4) - 3)\operatorname{tr}(\mathbf{A}^{\circ 2})\right)^2} \to 0. \tag{6}$$

It can be seen that (6) is much weaker than (5). For example, if $a_{i,j} = 1$, $i = 1, \dots, n$, $j = 1, \dots, n$ and $\mathrm{E}(\xi_1^4) = 3$, then the condition (6) holds but the condition (5) does not hold.

We now apply Theorem 1 to approximate the null distribution of the proposed statistic T. For $n \times n$ matrix $\mathbf{A}$ and real number τ, let $F(x; \mathbf{A}, \tau)$ be the cumulative distribution function of $\tau \sum_{i=1}^n a_{i,i}\check{z}_i + 2\sum_{1\le i<j\le n} a_{i,j} z_i z_j$, where $z_1, \dots, z_n, \check{z}_1, \dots, \check{z}_n$ are iid $\mathcal{N}(0,1)$ random variables.

Under the null hypothesis, Theorem 1 implies that

$$\begin{aligned}
&\Pr(T > x) \\
=&\Pr\left((\sqrt{\phi}\boldsymbol{\epsilon})^\top \left(-\tilde{\mathbf{U}}_a(\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1}\tilde{\mathbf{U}}_a^\top - x\tilde{\mathbf{U}}_a\tilde{\mathbf{U}}_a^\top\right)(\sqrt{\phi}\boldsymbol{\epsilon}) > 0\right) \\
\approx&1 - F\left(\operatorname{tr}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1}\right) + (n-q)x; -\tilde{\mathbf{U}}_a(\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1}\tilde{\mathbf{U}}_a^\top - x\tilde{\mathbf{U}}_a\tilde{\mathbf{U}}_a^\top, \sqrt{\phi^2\,\mathrm{E}(\epsilon_1^4) - 1}\right).
\end{aligned}$$

Ideally, we should find a consistent estimator $\hat{\tau}$ of $\sqrt{\phi^2 \operatorname{E}(\epsilon_1^4) - 1}$ and solve x from the equation

$$F\left(\operatorname{tr}(\mathbf{X}_b^* \mathbf{X}_b^{*\top})^{-1} + (n-q)x; -\tilde{\mathbf{U}}_a(\mathbf{X}_b^* \mathbf{X}_b^{*\top})^{-1}\tilde{\mathbf{U}}_a^\top - x\tilde{\mathbf{U}}_a\tilde{\mathbf{U}}_a^\top, \hat{\tau}\right) = 1 - \alpha.$$

However, solving this equation is not an easy task. For ease of implementation, we propose a one step iteration algorithm. We set the start point as

$$x^{(0)} = -\frac{\operatorname{tr}(\mathbf{X}_b^* \mathbf{X}_b^{*\top})^{-1}}{n-q},$$

which is chosen such that $\operatorname{tr}(-\tilde{\mathbf{U}}_a(\mathbf{X}_b^* \mathbf{X}_b^{*\top})^{-1}\tilde{\mathbf{U}}_a^\top - x^{(0)}\tilde{\mathbf{U}}_a\tilde{\mathbf{U}}_a^\top) = 0$. Let

$$x^{(1)} = \frac{F^{-1}\left(1-\alpha; -\tilde{\mathbf{U}}_a(\mathbf{X}_b^* \mathbf{X}_b^{*\top})^{-1}\tilde{\mathbf{U}}_a^\top - x^{(0)}\tilde{\mathbf{U}}_a\tilde{\mathbf{U}}_a^\top, \hat{\tau}\right) - \operatorname{tr}(\mathbf{X}_b^* \mathbf{X}_b^{*\top})^{-1}}{n-q}.$$

As noted in Remark 1, τ^2 should be a consistent estimator of $\phi^2 \operatorname{E}(\epsilon_1^4) - 1$ under the null hypothesis.

Theorem 2. *Suppose Assumption 1 holds. Furthermore, suppose the distribution of ϵ_1 is symmetric about 0 and has finite eighth moments. Suppose $\operatorname{E}\epsilon_1^4 > \phi^{-2}$. Let*

$$\mathbf{A} = -\tilde{\mathbf{U}}_a(\mathbf{X}_b^* \mathbf{X}_b^{*\top})^{-1}\tilde{\mathbf{U}}_a^\top - x^{(0)}\tilde{\mathbf{U}}_a\tilde{\mathbf{U}}_a^\top = -\tilde{\mathbf{U}}_a(\mathbf{X}_b^* \mathbf{X}_b^{*\top})^{-1}\tilde{\mathbf{U}}_a^\top + \frac{\operatorname{tr}\left((\mathbf{X}_b^* \mathbf{X}_b^{*\top})^{-1}\right)}{n-q}\tilde{\mathbf{U}}_a\tilde{\mathbf{U}}_a^\top.$$

Suppose as $n \to \infty$,

$$\frac{\max_{i,j} a_{i,j}^2}{\operatorname{tr}(\mathbf{A})^2} \to 0. \tag{7}$$

Let $\hat{\tau}^2$ be an consistent estimator of $\phi^2 \operatorname{E}(\epsilon_1^4) - 1$ based on $\mathbf{X}$, $\mathbf{y}$. Then

$$\Pr\left(T > x^{(1)}\right) \to \alpha.$$

Bai et al. Bai et al. (2018) gave a consistent estimator of $\sigma^{-4}\operatorname{E}(\epsilon_1^4) - 1$ based on the standardized residuals. Here we use a slightly different estimator which is based on the ordinary least squares residuals $\tilde{\boldsymbol{\epsilon}} = (\tilde{\epsilon}_1, \ldots, \tilde{\epsilon}_n)^\top = \tilde{\mathbf{P}}_a\mathbf{y}$. From Bai et al. (2018), Theorem 2.1,

$$\begin{aligned}
&\operatorname{E}\left(\tilde{\boldsymbol{\epsilon}}^\top\tilde{\mathbf{P}}_a\tilde{\boldsymbol{\epsilon}}\right) = (n-q)\sigma^2, \\
&\operatorname{E}\left(\sum_{i=1}^n \tilde{\epsilon}_i^4\right) = 3\sigma^4\operatorname{tr}(\tilde{\mathbf{P}}_a^{\circ 2}) + \left(\operatorname{E}(\epsilon_1^4) - 3\sigma^4\right)\operatorname{tr}\left(\tilde{\mathbf{P}}_a^{\circ 2}\right)^2.
\end{aligned}$$

Then a moment estimator of $\sigma^{-4}\operatorname{E}(\epsilon_1^4) - 1$ is

$$\hat{\tau}^2 = \frac{\dfrac{(n-q)^2\sum_{i=1}^n \tilde{\epsilon}_i^4}{\left(\tilde{\boldsymbol{\epsilon}}^\top\tilde{\mathbf{P}}_a\tilde{\boldsymbol{\epsilon}}\right)^2} - 3\operatorname{tr}(\tilde{\mathbf{P}}_a^{\circ 2})}{\operatorname{tr}\left(\tilde{\mathbf{P}}_a^{\circ 2}\right)^2} + 2.$$

Proposition 2. *Suppose the assumptions of Theorem 2 hold. Furthermore, suppose $q/n \to 0$. Then under the null hypothesis, $\hat{\tau}^2 \xrightarrow{P} \sigma^{-4}\operatorname{E}(\epsilon_1^4) - 1$.*

We reject the null hypothesis if

$$T > x^{(1)}.$$

This test procedure is asymptotically exact of size α under the conditions of Theorem 2 and Proposition 2.

2.3 Power analysis

In this section, we investigate the asymptotic power of the proposed test procedure as well as the test of Goeman et al. Goeman et al. (2006). To make the expression of the asymptotic power functions tractable, we shall assume further conditions so that the test statistics are asymptotically normally distributed. Also, $\boldsymbol{\epsilon}$ is assumed to be normally distributed so that we can can obtain the global power function rather than only local power function.

To derive the asymptotic power of the proposed test and the test of Goeman et al. Goeman et al. (2006) simultaneously, we consider the general statistic $\mathbf{y}^{*\top}(\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k\mathbf{y}^*/\mathbf{y}^{*\top}\mathbf{y}^*$. In fact, the proposed test statistic corresponds to $k = -1$ while the test statistic of Goeman et al. Goeman et al. (2006) corresponds to $k = 1$. Note that for any $x \in \mathbb{R}$,

$$\Pr\left(\frac{\mathbf{y}^{*\top}(\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k\mathbf{y}^*}{\mathbf{y}^{*\top}\mathbf{y}^*} \le x\right) = \Pr\left(\mathbf{y}^{*\top}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k - x\mathbf{I}_{n-q}\right)\mathbf{y}^* \le 0\right).$$

Hence the asymptotic behavior of noncentral quadratic form will play a key role in our investigation. We have the following proposition.

Proposition 3. *Let $Z = (z_1, \ldots, z_n)^\top$, where z_i's are iid $\mathcal{N}(0,1)$ random variables. Let $\mathbf{A}$ be an $n \times n$ symmetric matrix with elements $a_{i,j}$. Let $\mathbf{b} = (b_1, \ldots, b_n)^\top$ be an n dimensional vector. If $\operatorname{tr}(\mathbf{A}^4)/\operatorname{tr}^2(\mathbf{A}^2) \to 0$, then*

$$\frac{Z^\top\mathbf{A}Z + b^\top Z - \operatorname{tr}(\mathbf{A})}{\sqrt{2\operatorname{tr}(\mathbf{A}^2) + \|\mathbf{b}\|^2}} \rightsquigarrow \mathcal{N}(0,1).$$

Remark 3. Proposition 3 does not impose any condition on $\mathbf{b}$. This allows us to give the global asymptotic power function of tests. As the cost of this flexibility, we have to make the normal assumption.

Now we investigate the asymptotic behavior of $\mathbf{y}^{*\top}(\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k\mathbf{y}^*/\mathbf{y}^{*\top}\mathbf{y}^*$. Let $w_i = (\mathbf{V}_b^{*\top}\boldsymbol{\beta}_b)_i$ be the coordinate of β_b along the ith principal component direction of $\mathbf{X}_b^{*\top}\mathbf{X}_b^*$, $i = 1, \ldots, n-q$. It turns out that many quantities involved can be conveniently represented as the expectations with respect to I, a random variable uniformly distributed on $\{1, \ldots, n-q\}$. For example, $\operatorname{tr}(\mathbf{X}_b^*\mathbf{X}_b^{*\top})^k/(n-q) = \operatorname{E}(\gamma_I^k)$.

Theorem 3. *Suppose model (1) holds with $\boldsymbol{\epsilon} \sim \mathcal{N}_n(0, \phi^{-1}\mathbf{I}_n)$. Let $k \neq 0$ be a fixed number. Suppose as $n \to \infty$, $n - q \to \infty$ and*

$$\frac{\max_{1 \leq i \leq n-q} \left(\gamma_i^k - \mathrm{E}(\gamma_I^k)\right)^2}{(n-q)\operatorname{Var}(\gamma_I^k)} \to 0. \tag{8}$$

Then for any $x \in \mathbb{R}$,

$$\begin{aligned}
&\Pr\left(\frac{\mathbf{y}^{*\top}\left(\mathbf{X}_b^*\mathbf{X}_b^{*\top}\right)^k\mathbf{y}^*}{\mathbf{y}^{*\top}\mathbf{y}^*} \leq \mathrm{E}(\gamma_I^k) + \sqrt{\frac{2\operatorname{Var}\left(\gamma_I^k\right)}{n-q}}x\right) \\
=&\Phi\left(\frac{\left(\mathrm{E}(\gamma_I w_I^2) + \phi^{-1}\right)\sqrt{2\operatorname{Var}\left(\gamma_I^k\right)}x - \sqrt{n-q}\operatorname{Cov}\left(\gamma_I^k, \gamma_I w_I^2\right)}{\sqrt{2\phi^{-2}\operatorname{Var}(\gamma_I^k) + 4\phi^{-1}\mathrm{E}\left[\left(\gamma_I^k - \mathrm{E}(\gamma_I^k) - \sqrt{\frac{2\operatorname{Var}(\gamma_I^k)}{n-q}}x\right)^2\gamma_I w_I^2\right]}}\right) + o(1).
\end{aligned}$$

Under the conditions of Theorem 3, the proposed test should reject the null hypothesis when

$$\frac{\mathbf{y}^{*\top}\left(\mathbf{X}_b^*\mathbf{X}_b^{*\top}\right)^{-1}\mathbf{y}^*}{\mathbf{y}^{*\top}\mathbf{y}^*} \leq \mathrm{E}(\gamma_I^{-1}) + \sqrt{\frac{2\operatorname{Var}\left(\gamma_I^{-1}\right)}{n-q}}\Phi^{-1}(\alpha),$$

and the asymptotic power function of the proposed test is

$$\Phi\left(\frac{\left(\mathrm{E}(\gamma_I w_I^2) + \phi^{-1}\right)\sqrt{2\operatorname{Var}\left(\gamma_I^{-1}\right)}\Phi^{-1}(\alpha) + \sqrt{n-q}\operatorname{Cov}\left(-\gamma_I^{-1}, \gamma_I w_I^2\right)}{\sqrt{2\phi^{-2}\operatorname{Var}(\gamma_I^{-1}) + 4\phi^{-1}\mathrm{E}\left[\left(\gamma_I^{-1} - \mathrm{E}(\gamma_I^{-1}) - \sqrt{\frac{2\operatorname{Var}(\gamma_I^{-1})}{n-q}}\Phi^{-1}(\alpha)\right)^2\gamma_I w_I^2\right]}}\right). \tag{9}$$

On the other hand, the test of Goeman et al. Goeman et al. (2006) should reject the null hypothesis when

$$\frac{\mathbf{y}^{*\top}\mathbf{X}_b^*\mathbf{X}_b^{*\top}\mathbf{y}^*}{\mathbf{y}^{*\top}\mathbf{y}^*} > \mathrm{E}(\gamma_I) + \sqrt{\frac{2\operatorname{Var}\left(\gamma_I\right)}{n-q}}\Phi^{-1}(1-\alpha),$$

and the asymptotic power function of their test is

$$\Phi\left(\frac{\left(\mathrm{E}(\gamma_I w_I^2) + \phi^{-1}\right)\sqrt{2\operatorname{Var}\left(\gamma_I\right)}\Phi^{-1}(\alpha) + \sqrt{n-q}\operatorname{Cov}\left(\gamma_I, \gamma_I w_I^2\right)}{\sqrt{2\phi^{-2}\operatorname{Var}(\gamma_I) + 4\phi^{-1}\mathrm{E}\left[\left(\gamma_I - \mathrm{E}(\gamma_I) + \sqrt{\frac{2\operatorname{Var}(\gamma_I)}{n-q}}\Phi^{-1}(\alpha)\right)^2\gamma_I w_I^2\right]}}\right). \tag{10}$$

It can be seen from (9) and (10) that the power of the proposed test mainly depends on $\operatorname{Cov}(-\gamma_I^{-1}, \gamma_I w_I^2)$ while the power of the test of Goeman et al. Goeman et al. (2006) mainly depends on $\mathrm{Cov}(\gamma_I, \gamma_I w_I^2)$. Unfortunately, neither of these two quantities is positive definite. This fact is not surprising in view of Proposition 1. Nevertheless, $\mathrm{Cov}(-\gamma_I^{-1}, \gamma_I w_I^2)$ and $\mathrm{Cov}(\gamma_I, \gamma_I w_I^2)$ are positive definite if $\boldsymbol{\beta}_b$ is restricted in certain subspaces of $\mathbb{R}^p$.

Let d_1 be the maximum i such that $\gamma_i^{-1} < \mathrm{E}(\gamma_I^{-1})$. Let d_2 be the maximum i such that $\gamma_i > \mathrm{E}(\gamma_I)$. Then it can be seen that

$$\mathrm{Cov}(-\gamma_I^{-1}, \gamma_I w_I^2) = \frac{1}{n-q} \sum_{i=1}^{n-q} \left(\mathrm{E}(\gamma_I^{-1}) - \gamma_i^{-1}\right) \gamma_i w_i^2$$

is positive definite if $w_{d_1+1} = \cdots = w_{n-q} = 0$, while

$$\mathrm{Cov}(\gamma_I, \gamma_I w_I^2) = \frac{1}{n-q} \sum_{i=1}^{n-q} (\gamma_i - \mathrm{E}\,\gamma_I)\, \gamma_i w_i^2$$

is positive definite if $w_{d_2+1} = \cdots = w_{n-q} = 0$. In other words, $\mathrm{Cov}(-\gamma_I^{-1}, \gamma_I w_I^2)$ is positive definite if $\boldsymbol{\beta}_b$ belongs to the rank d_1 principal component subspace of $\mathbf{X}_b^{*\top}\mathbf{X}_b^*$, and $\mathrm{Cov}(\gamma_I, \gamma_I w_I^2)$ is positive definite if $\boldsymbol{\beta}_b$ belongs to the rank d_2 principal component subspace of $\mathbf{X}_b^{*\top}\mathbf{X}_b^*$. Note that $\gamma_i > \mathrm{E}(\gamma_I)$ implies that $\gamma_i^{-1} < (\mathrm{E}(\gamma_I))^{-1} \leq \mathrm{E}(\gamma_I^{-1})$, where the last inequality follows from Jensen's inequality. It follows that $d_1 \geq d_2$. Hence the proposed test can always detect the signals from more directions than the test of Goeman et al. Goeman et al. (2006).

Now we consider a practical scenario where the proposed test is more favorable. Suppose $\mathbf{X}_b$ is a random design matrix and the rows of $\mathbf{X}_b$ are iid random vectors. Let $X_{b,i} \in \mathbb{R}^p$ be the transpose of the ith row of $\mathbf{X}_b$, $i = 1, \ldots n$. Often in practice, the predictors are affected by a few common factors. Specifically, we assume $X_{b,i}$ has structre $X_{b,i} = \mathbf{B}\mathbf{f}_i + u_i$, $\mathbf{f}_i \in \mathbb{R}^K$ is the vector of common factors, K is factor number, $\mathbf{B}$ is a $p \times K$ loading matrix, and $u_i \in \mathbb{R}^p$ is the vector of idiosyncratic components, $i = 1, \ldots, n$. In this case, $\gamma_1, \ldots, \gamma_K$ are typically much larger than $\gamma_{K+1}, \ldots, \gamma_{n-q}$ (See, e.g., Fan et al. (2013)). Note that $d_2 = K$ for large $\gamma_1, \ldots, \gamma_K$. In this case, the test of Goeman et al. Goeman et al. (2006) can only detect the signals from the rank K principal component subspace of $\mathbf{X}_b^{*\top}\mathbf{X}_b^*$. In some situations, however, the effect of $\mathbf{X}_b$ on $\mathbf{y}$ is mainly made by the idiosyncratic components. In this case, $\boldsymbol{\beta}_b$ is almost orthogonal to the rank K principal component subspace of $\mathbf{X}_b^{*\top}\mathbf{X}_b^*$, and consequently, the test of Goeman et al. Goeman et al. (2006) will have poor behavior. On the other hand, the large eigenvalues $\gamma_1, \ldots, \gamma_K$ has little effect on d_1. Hence the proposed test is expected to have good performance.

3 Numerical results

In this section, we conduct simulations to examine the empirical size and power of the proposed test (abbreviated as NEW) and compare it with the global test proposed by Goeman et al. Goeman et al. (2006) (abbreviated as GT) and the EigenPrism test prospoed by Janson et al. Janson et al. (2016) (abbreviated as EP). Throughout our simulations, we take $q = 10$, $p = 1000$, $\alpha = 0.05$ and the empirical results are obtained from 2500 replications.

In our simulations, two distributions of ϵ_1 are considered, namely the chi-squared distribution $\epsilon_1 \sim (\chi^2(4) - 4)/\sqrt{8}$ and the Student's t distribution $\epsilon_1 \sim t_9$. Define the signal to noise ratio (SNR) as $\mathrm{SNR} = \sqrt{(n-q)\operatorname{Var}(\gamma_I)\phi\|\boldsymbol{\beta}_b\|^2/p}$. We consider two structures of $\boldsymbol{\beta}_b$: dense $\boldsymbol{\beta}_b$ and sparse $\boldsymbol{\beta}_b$. In dense $\boldsymbol{\beta}_b$ setting, the coordinates of $\boldsymbol{\beta}_b$ are independently generated from the uniform distribution $U(-c, c)$ where c is selected to reach certain SNR. In sparse $\boldsymbol{\beta}_b$ setting, we randomly select 5% of the coordinates of $\boldsymbol{\beta}_b$ to be non-zero and the non-zero coordinates are independently generated from $U(-c, c)$ where c is selected to reach certain SNR. The design matrices $\mathbf{X}_a$ and $\mathbf{X}_b$ are randomly generated beforehand and are fixed during each simulation. In all simulations, the elements of $\mathbf{X}_a$ are iid from standard normal distribution. We consider three models of $\mathbf{X}_b$, as follows.

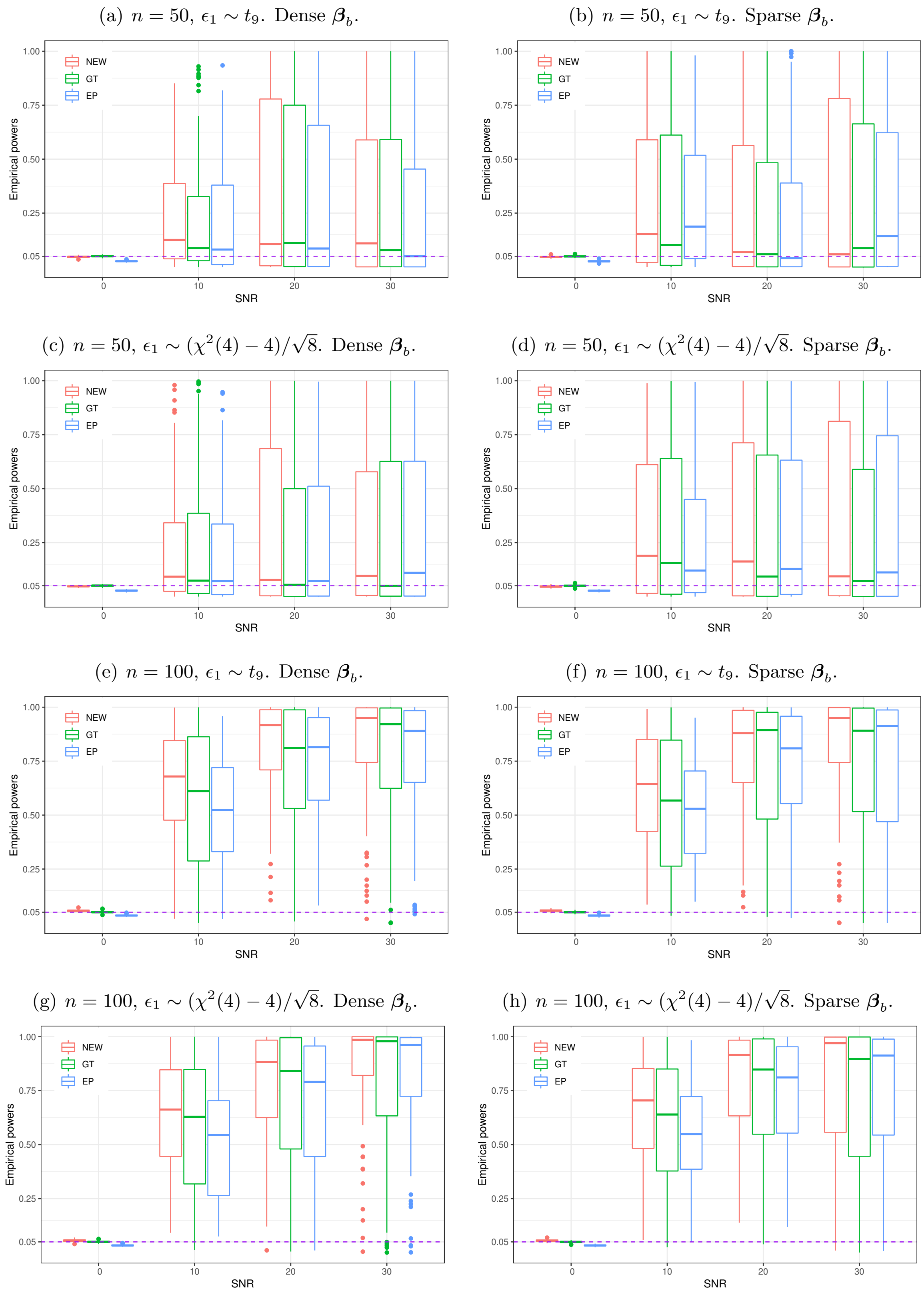
- Model I: the rows of $\mathbf{X}_b$ are iid from $\mathcal{N}(0, \boldsymbol{\Sigma}_b)$, where $\boldsymbol{\Sigma}_b = \Gamma\Gamma^\top$ and Γ is a $p \times p$ matrix whose elements are iid generated from $\mathcal{N}(0, 1)$.
- Model II: the rows of $\mathbf{X}_b$ are iid from $\mathcal{N}(0, \boldsymbol{\Sigma}_b)$, where $(\boldsymbol{\Sigma}_b)_{i,j} = 0.1$ if $i \neq j$ and $(\boldsymbol{\Sigma}_b)_{i,j} = 1$ otherwise.
- Model III: $\mathbf{X}_b$ is generated as Model II but the observed design matrix is $\mathbf{F}\mathbf{B}^\top + \mathbf{X}_b$, where $\mathbf{F}$ is an $n \times 2$ random matrix with iid $\mathcal{N}(0, 1)$ entries and $\mathbf{B}$ is a $p \times 2$ loading matrix with iid $\mathcal{N}(0, 1)$ entries.

The simulation results are illustrated in Figures 1-3. From the results, it can be seen that all three tests can well maintain the significant level, although the EP test is a little conservative in some cases. Note that for $n = 50$, the powers of tests typically do not converge to 1 as SNR increases. This phenomenon, which is not surprising in view of Proposition 1, is eased when the sample size rises to 100. In all three simulations, the proposed test shows relatively good power behavior compared with its two competitors. Figure 3 shows that under Model III, the GT test has negligible power while the proposed test has reasonable power. This verifies the argument made at the end of Section 2.3.

4 Conclusions

We have proposed a Bayesian-motivated test statistic for high-dimensional linear model with fixed design. We proposed an approximation of the null distribution of the proposed test statistic which is then used to determine the critical value of the test statistic. Under weak conditions, we proved the proposed test procedure is asymptotically level α. Under certain conditions, we also derived the asymptotic power function of the proposed test.

The methodology of Goeman et al. Goeman et al. (2006) can also be applied in generalized linear models. However, it is not obvious how to generalize the proposed test and the test of Janson et al. Janson et al. (2016) to generalized linear models. More fundamentally, it is unclear if results like Proposition 1 hold for generalized linear model.



(a) $n = 50$, $\epsilon_1 \sim t_9$. Dense $\boldsymbol{\beta}_b$.

(b) $n = 50$, $\epsilon_1 \sim t_9$. Sparse $\boldsymbol{\beta}_b$.

(c) $n = 50$, $\epsilon_1 \sim (\chi^2(4) - 4)/\sqrt{8}$. Dense $\boldsymbol{\beta}_b$.

(d) $n = 50$, $\epsilon_1 \sim (\chi^2(4) - 4)/\sqrt{8}$. Sparse $\boldsymbol{\beta}_b$.

(e) $n = 100$, $\epsilon_1 \sim t_9$. Dense $\boldsymbol{\beta}_b$.

(f) $n = 100$, $\epsilon_1 \sim t_9$. Sparse $\boldsymbol{\beta}_b$.

(g) $n = 100$, $\epsilon_1 \sim (\chi^2(4) - 4)/\sqrt{8}$. Dense $\boldsymbol{\beta}_b$.

(h) $n = 100$, $\epsilon_1 \sim (\chi^2(4) - 4)/\sqrt{8}$. Sparse $\boldsymbol{\beta}_b$.

Figure 1: Box plots of the empirical powers based on 100 independently generated $\boldsymbol{\beta}_b$. $\mathbf{X}_b$ is generated by Model I.

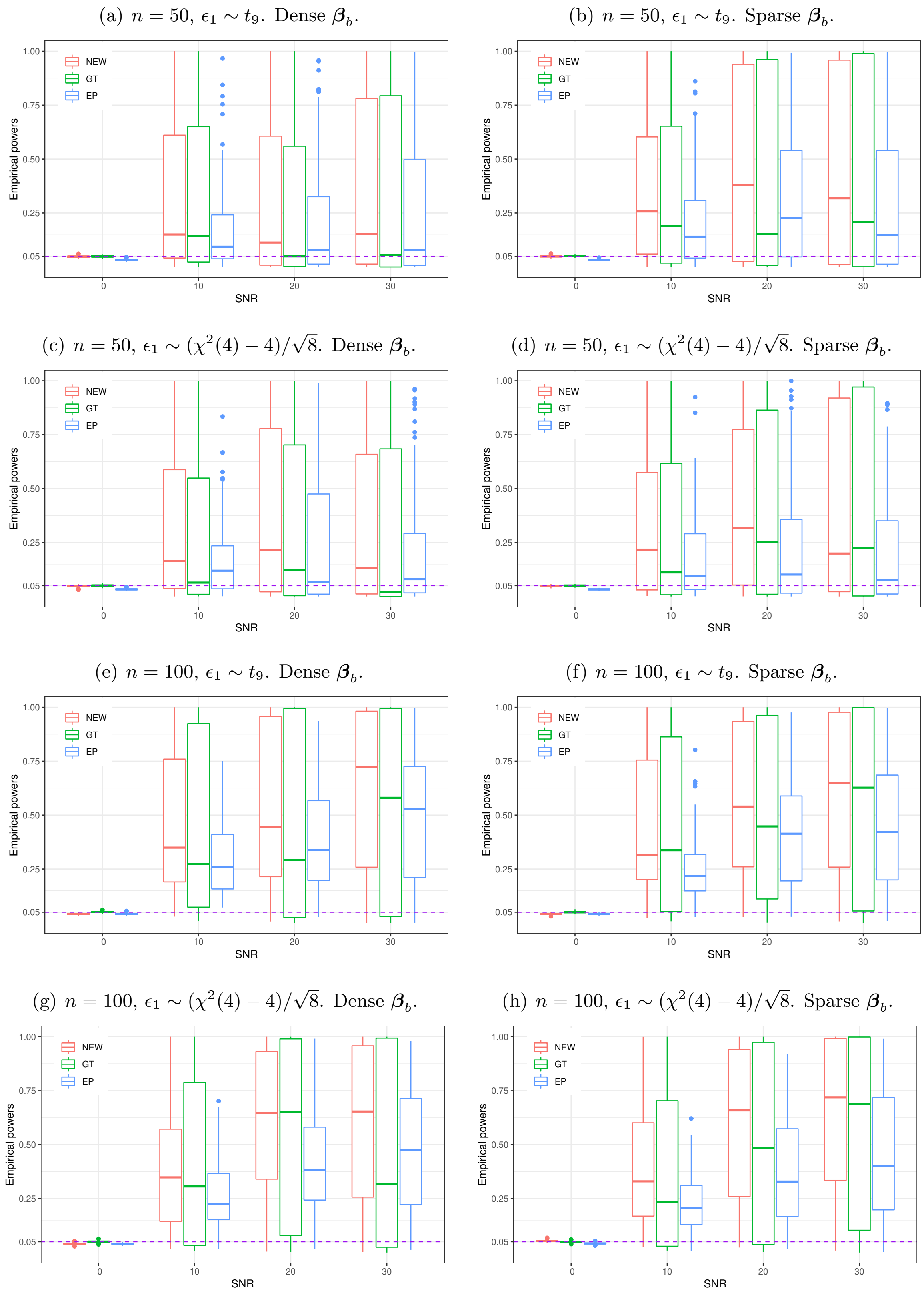
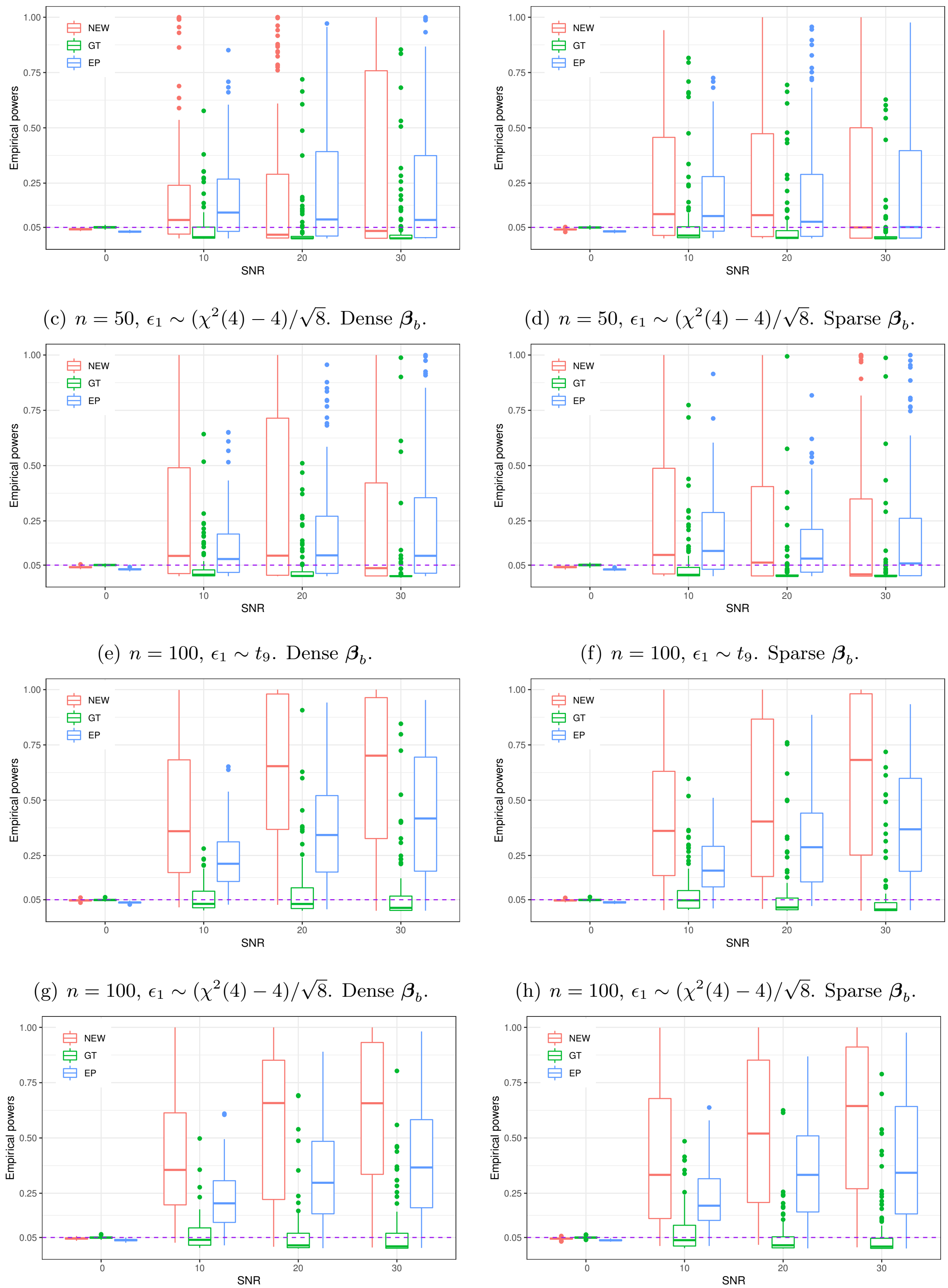


Figure 2: Box plots of the empirical powers based on 100 independently generated $\boldsymbol{\beta}_b$. $\mathbf{X}_b$ is generated by Model II.



(a) $n = 50$, $\epsilon_1 \sim t_9$. Dense $\boldsymbol{\beta}_b$.

(b) $n = 50$, $\epsilon_1 \sim t_9$. Sparse $\boldsymbol{\beta}_b$.

(c) $n = 50$, $\epsilon_1 \sim (\chi^2(4) - 4)/\sqrt{8}$. Dense $\boldsymbol{\beta}_b$.

(d) $n = 50$, $\epsilon_1 \sim (\chi^2(4) - 4)/\sqrt{8}$. Sparse $\boldsymbol{\beta}_b$.

(e) $n = 100$, $\epsilon_1 \sim t_9$. Dense $\boldsymbol{\beta}_b$.

(f) $n = 100$, $\epsilon_1 \sim t_9$. Sparse $\boldsymbol{\beta}_b$.

(g) $n = 100$, $\epsilon_1 \sim (\chi^2(4) - 4)/\sqrt{8}$. Dense $\boldsymbol{\beta}_b$.

(h) $n = 100$, $\epsilon_1 \sim (\chi^2(4) - 4)/\sqrt{8}$. Sparse $\boldsymbol{\beta}_b$.

Figure 3: Box plots of the empirical powers based on 100 independently generated $\boldsymbol{\beta}_b$. $\mathbf{X}_b$ is generated by Model III.

In Theorem 1, we assumed that the distribution of ϵ_1 is symmetric about 0. This condition is also assumed by Bai et al. Bai et al. (2018) in the study of central limit theorem of quadratic form. Without this condition, our proof of Theorem 1 is not valid. We think it will be an interesting and useful work to relax the symmetric condition in Theorem 1.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant No. 11471035.

Appendix

Lemma 1. *Under the assumptions of Proposition 1, if there exists a Borel set $G \subset \mathbb{R}^n$ and a number $M \geq 0$ such that*

$$\int_{\mathbb{R}^n} \varphi(\mathbf{y}) d\mathcal{N}_n(\mu, \phi^{-1}\mathbf{I}_n)(\mathbf{y})\, \mathrm{d}\mathbf{y} \geq \alpha \quad \textit{for all } \mu \in G \textit{ and } \phi > M,$$

then $\lambda(\{\mathbf{y} : \phi(\mathbf{y}) < \alpha, \mathbf{y} \in G\}) = 0$.

Proof. We prove the claim by contradiction. Suppose $\lambda(\{\mathbf{y} : \varphi(\mathbf{y}) < \alpha, \mathbf{y} \in G\}) > 0$. Then there exists a sufficiently small $0 < \eta < \alpha$, such that $\lambda(\{\mathbf{y} : \varphi(\mathbf{y}) < \alpha - \eta, \mathbf{y} \in G\}) > 0$. We denote $E := \{\mathbf{y} \in \mathbb{R}^n : \varphi(\mathbf{y}) < \alpha - \eta, \mathbf{y} \in G\}$. From Lebesgue density theorem (Cohn, 2013, Corollary 6.2.6), there exists a point $z \in E$, such that, for any $\varepsilon > 0$ there is a $\delta_\varepsilon > 0$ such that for any $0 < \delta' < \delta_\varepsilon$,

$$\left| \frac{\lambda(E^\complement \cap C_{\delta'})}{\lambda(C_{\delta'})} \right| < \varepsilon,$$

where $C_{\delta'} = \prod_{i=1}^n [z_i - \delta', z_i + \delta']$. We put

$$\varepsilon = \left(\frac{\sqrt{\pi}}{\sqrt{2}\Phi^{-1}\left(1 - \frac{\eta}{6n}\right)} \right)^n \frac{\eta}{3},$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal random variable. Then for any $\phi > M$ and $0 < \delta' < \delta_\varepsilon$,

$$
\begin{aligned}
\alpha \leq& \int_{\mathbb{R}^n} \varphi(\mathbf{y}) d\mathcal{N}_n(z, \phi^{-1}\mathbf{I}_n)(\mathbf{y})\, \mathrm{d}\mathbf{y} \\
=& \int_{E\cap C_{\delta'}} \varphi(\mathbf{y}) d\mathcal{N}_n(z, \phi^{-1}\mathbf{I}_n)(\mathbf{y})\, \mathrm{d}\mathbf{y} + \int_{E^{\complement}\cap C_{\delta'}} \varphi(\mathbf{y}) d\mathcal{N}_n(z, \phi^{-1}\mathbf{I}_n)(\mathbf{y})\, \mathrm{d}\mathbf{y} \\
&+ \int_{C_{\delta'}^{\complement}} \varphi(\mathbf{y}) d\mathcal{N}_n(z, \phi^{-1}\mathbf{I}_n)(\mathbf{y})\, \mathrm{d}\mathbf{y} \\
\leq& \alpha - \eta + \int_{E^{\complement}\cap C_{\delta'}} d\mathcal{N}_n(z, \phi^{-1}\mathbf{I}_n)(\mathbf{y})\, \mathrm{d}\mathbf{y} + \int_{C_{\delta'}^{\complement}} d\mathcal{N}_n(z, \phi^{-1}\mathbf{I}_n)(\mathbf{y})\, \mathrm{d}\mathbf{y} \\
\leq& \alpha - \eta + \left(\frac{\phi}{2\pi}\right)^{n/2} \lambda(E^{\complement}\cap C_{\delta'}) + 2n\left(1 - \Phi(\sqrt{\phi}\delta')\right) \\
\leq& \alpha - \eta + \left(\frac{\phi}{2\pi}\right)^{n/2} \varepsilon(2\delta')^n + 2n\left(1 - \Phi(\sqrt{\phi}\delta')\right) \\
=& \alpha - \eta + \left(\frac{\sqrt{\phi}\delta'}{\Phi^{-1}\left(1-\frac{\eta}{6n}\right)}\right)^n \frac{\eta}{3} + 2n\left(1 - \Phi(\sqrt{\phi}\delta')\right).
\end{aligned}
$$

In the last inequality, we put δ' small enough such that

$$
\left(\frac{\Phi^{-1}\left(1-\frac{\eta}{6n}\right)}{\delta'}\right)^2 > M,
$$

and put

$$
\phi = \left(\frac{\Phi^{-1}\left(1-\frac{\eta}{6n}\right)}{\delta'}\right)^2.
$$

Then we obtain the contradiction $\alpha \leq \alpha - (2/3)\eta$. This completes the proof.

$\square$

Proof of Proposition 1. To prove the claim (a), note that $\varphi(\mathbf{y})$ is unbiased if and only if

$$
\int_{\mathbb{R}^n} \varphi(\mathbf{y}) d\mathcal{N}_n(\mu, \phi^{-1}\mathbf{I}_n)(\mathbf{y})\, \mathrm{d}\mathbf{y} \geq \alpha \quad \text{for all } \mu \in \mathbb{R}^n \text{ and } \phi > 0.
$$

Then Lemma 1 implies that $\varphi(\mathbf{y}) \geq \alpha$, a.s. λ. On the other hand, since $\varphi(\mathbf{y})$ is a level α test, for every $\phi > 0$,

$$
\int_{\mathbb{R}^n} [\varphi(\mathbf{y}) - \alpha] d\mathcal{N}_n(0, \phi^{-1}\mathbf{I}_n)(\mathbf{y})\, \mathrm{d}\mathbf{y} \leq 0. \tag{11}
$$

Note that the integrand of (11) is nonnegative. Hence $\varphi(\mathbf{y}) - \alpha = 0$ a.s. λ. This proves the claim (a).

Now we prove the claim (b) by contradiction. Suppose

$$
\int_{\mathbb{R}^n} \varphi(\mathbf{y}) d\mathcal{N}_n\left(\mathbf{X}_a\boldsymbol{\beta}_a + \mathbf{X}_b\boldsymbol{\beta}_b, \phi^{-1}\mathbf{I}_n\right)(\mathbf{y})\, \mathrm{d}\mathbf{y} \geq \alpha
$$

for every $\boldsymbol{\beta}_a \in \mathbb{R}^q$, $\boldsymbol{\beta}_b \in \mathbb{R}^p$, $\phi > 0$ satisfying $\phi\boldsymbol{\beta}_b^\top\mathbf{X}_b^\top\tilde{\mathbf{P}}_a\mathbf{X}_b\boldsymbol{\beta}_b > M$. Note that

$$
\left\{(\boldsymbol{\beta}_b^\top, \phi)^\top : \boldsymbol{\beta}_b^\top\mathbf{X}_b^\top\tilde{\mathbf{P}}_a\mathbf{X}_b\boldsymbol{\beta}_b > \sqrt{M}, \phi > \sqrt{M}\right\} \subset \left\{(\boldsymbol{\beta}_b^\top, \phi)^\top : \phi\boldsymbol{\beta}_b^\top\mathbf{X}_b^\top\tilde{\mathbf{P}}_a\mathbf{X}_b\boldsymbol{\beta}_b > M\right\}.
$$

Then Lemma 1 implies that $\varphi(\mathbf{y})\mathbf{1}_G(\mathbf{y}) \geq \alpha \mathbf{1}_G(\mathbf{y})$ a.e. λ, where

$$G = \left\{ \mathbf{X}_a \boldsymbol{\beta}_a + \mathbf{X}_b \boldsymbol{\beta}_b : \boldsymbol{\beta}_a \in \mathbb{R}^q, \boldsymbol{\beta}_b^\top \mathbf{X}_b^\top \tilde{\mathbf{P}}_a \mathbf{X}_b \boldsymbol{\beta}_b > \sqrt{M} \right\}.$$

It can be easily seen that

$$G = \left\{ w + z : w \in \operatorname{span}(\mathbf{X}_a), z \in \operatorname{span}(\mathbf{X}_a)^\perp, \|z\|^2 > \sqrt{M} \right\},$$

where $\operatorname{span}(\mathbf{X}_a)$ is the linear span of the columns of $\mathbf{X}_a$ and $\operatorname{span}(\mathbf{X}_a)^\perp$ is the orthogonal complement of $\operatorname{span}(\mathbf{X}_a)$. Since $\varphi(\mathbf{y})$ is nonrandom and $\alpha > 0$, we have $\varphi(\mathbf{y})\mathbf{1}_G(\mathbf{y}) \geq \mathbf{1}_G(\mathbf{y})$ a.e. λ. Also note that $\varphi(\mathbf{y})$ is a level α test, then for every $\phi > 0$, we have

$$\begin{aligned}
\alpha \geq & \int_{\mathbb{R}^n} \varphi(\mathbf{y}) d\mathcal{N}_n(0, \phi^{-1}\mathbf{I}_n)(\mathbf{y})\, \mathrm{d}\mathbf{y} \\
\geq & \int_G d\mathcal{N}_n(0, \phi^{-1}\mathbf{I}_n)(\mathbf{y}) \\
= & \Pr\left(R > \phi M^{*2}\right),
\end{aligned}$$

where R is a random variable with distribution $\chi^2(n-q)$. Thus, we obtain a contradiction by letting $\phi \to 0$ in the above inequality. This completes the proof.

□

Proof of Theorem 1. Let

$$\tilde{a}_{i,j} := \frac{a_{i,j}}{\sqrt{2\operatorname{tr}(\mathbf{A}^2) + (\mathrm{E}(\xi_1^4) - 3)\operatorname{tr}(\mathbf{A}^{\circ 2})}}.$$

Then

$$S = \sum_{i=1}^n \tilde{a}_{i,i}(\xi_i^2 - 1) + 2\sum_{1 \leq i < j \leq n} \tilde{a}_{i,j}\xi_i\xi_j, \quad S_\tau^* = \tau \sum_{i=1}^n \tilde{a}_{i,i}\check{z}_i + 2\sum_{1 \leq i < j \leq n} \tilde{a}_{i,j} z_i z_j.$$

For $l = 1, \ldots, n$, define

$$\begin{aligned}
S_l = & \sum_{i=1}^{l-1} \tilde{a}_{i,i}(\xi_i^2 - 1) + \tau \sum_{i=l+1}^n \tilde{a}_{i,i}\check{z}_i \\
& + 2\sum_{1 \leq i < j \leq l-1} \tilde{a}_{i,j}\xi_i\xi_j + 2\sum_{i=1}^{l-1}\sum_{j=l+1}^n \tilde{a}_{i,j}\xi_i z_j + 2\sum_{l+1 \leq i < j \leq n} \tilde{a}_{i,j} z_i z_j, \\
h_l = & \tilde{a}_{l,l}(\xi_l^2 - 1) + 2\sum_{i=1}^{l-1} \tilde{a}_{i,l}\xi_i\xi_l + 2\sum_{i=l+1}^n \tilde{a}_{i,l} z_i \xi_l, \\
g_l = & \tau \tilde{a}_{l,l}\check{z}_l + 2\sum_{i=1}^{l-1} \tilde{a}_{i,l}\xi_i z_l + 2\sum_{i=l+1}^n \tilde{a}_{i,l} z_i z_l.
\end{aligned}$$

It can be seen that for $l = 2, \ldots, n$, $S_{l-1} + h_{l-1} = S_l + g_l$, and $S = S_n + h_n$, $S_1 + g_1 = S_\tau^*$.

Thus, for any $f \in \mathscr{C}^4(\mathbb{R})$,

$$
\begin{aligned}
&|\mathrm{E}\, f\,(S) - \mathrm{E}\, f\,(S_\tau^*)| \\
=&|\mathrm{E}\, f(S_n + h_n) - \mathrm{E}\, f(S_1 + g_1)| \\
=&\left|\sum_{l=2}^{n} (\mathrm{E}\, f(S_l + h_l) - \mathrm{E}\, f(S_{l-1} + h_{l-1})) + \mathrm{E}\, f(S_1 + h_1) - \mathrm{E}\, f(S_1 + g_1)\right| \\
=&\left|\sum_{l=1}^{n} \mathrm{E}\, f(S_l + h_l) - \mathrm{E}\, f(S_l + g_l)\right|.
\end{aligned}
$$

Apply Taylor's theorem, for $l = 1, \ldots, n$,

$$
\begin{aligned}
f(S_l + h_l) =& f(S_l) + \sum_{k=1}^{3} \frac{1}{k!} h_l^k f^{(k)}(S_l) + \frac{1}{24} h_l^4 f^{(4)}(S_l + \theta_1 h_l), \\
f(S_l + g_l) =& f(S_l) + \sum_{k=1}^{3} \frac{1}{k!} g_l^k f^{(k)}(S_l) + \frac{1}{24} g_l^4 f^{(4)}(S_l + \theta_2 g_l),
\end{aligned}
$$

where $\theta_1, \theta_2 \in [0, 1]$. Thus,

$$
\begin{aligned}
&|\mathrm{E}\, f(S_l + h_l) - \mathrm{E}\, f(S_l + g_l)| \\
\leq &\left|\sum_{k=1}^{3} \frac{1}{k!} \mathrm{E}\, f^{(k)}(S_l)\, \mathrm{E}_l(h_l^k - g_l^k)\right| + \frac{1}{24} \|f''''\|_\infty \left(\mathrm{E}(h_l^4) + \mathrm{E}(g_l^4)\right),
\end{aligned}
$$

where E_l denotes taking expectation with respect to $\xi_l, z_l, \check{z}_l$. It is straightforward to show that

$$
\begin{aligned}
\mathrm{E}_l(h_l - g_l) &= 0, \\
\mathrm{E}_l(h_l^2 - g_l^2) &= \left(\mathrm{E}(\xi_1^4) - 1 - \tau^2\right) \tilde{a}_{l,l}^2, \\
\mathrm{E}_l(h_l^3 - g_l^3) &= \mathrm{E}(\xi_1^2 - 1)^3 \tilde{a}_{l,l}^3 + 12(\mathrm{E}(\xi_1^4) - 1)\tilde{a}_{l,l} \left(\sum_{i=1}^{l-1} \tilde{a}_{i,l}\xi_i + \sum_{i=l+1}^{n} \tilde{a}_{i,l} z_i\right)^2.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&|\mathrm{E}\, f(S_l + h_l) - \mathrm{E}\, f(S_l + g_l)| \\
\leq &\frac{1}{2} \|f^{(2)}\|_\infty \left|\mathrm{E}(\xi_1^4) - 1 - \tau^2\right| \tilde{a}_{l,l}^2 \\
&+ \frac{1}{6} \|f^{(3)}\|_\infty \left(\left|\mathrm{E}(\xi_1^2 - 1)^3\right| |\tilde{a}_{l,l}^3| + 12(\mathrm{E}(\xi_1^4) - 1)|\tilde{a}_{l,l}|\, \mathrm{E}\left(\sum_{i=1}^{l-1} \tilde{a}_{i,l}\xi_i + \sum_{i=l+1}^{n} \tilde{a}_{i,l} z_i\right)^2\right) \\
&+ \frac{1}{24} \|f^{(4)}\|_\infty \left(\mathrm{E}(h_l^4) + \mathrm{E}(g_l^4)\right) \\
\leq &\frac{\left|\mathrm{E}(\xi_1^4) - 1 - \tau^2\right|}{2} \|f^{(2)}\|_\infty \tilde{a}_{l,l}^2 + \frac{\max\left(\left|\mathrm{E}(\xi_1^2 - 1)^3\right|, 12(\mathrm{E}(\xi_1^4) - 1)\right)}{6} \|f^{(3)}\|_\infty |\tilde{a}_{l,l}| \sum_{i=1}^{n} \tilde{a}_{i,l}^2 \\
&+ \frac{1}{24} \|f^{(4)}\|_\infty \left(\mathrm{E}(h_l^4) + \mathrm{E}(g_l^4)\right).
\end{aligned}
\tag{12}
$$

Now we bound $\mathrm{E}(h_l^4)$ and $\mathrm{E}(g_l^4)$. By direct calculation,

$$\begin{aligned}
\mathrm{E}(h_l^4) =& \mathrm{E}(\xi_1^2-1)^4\tilde{a}_{l,l}^4 + 24\,\mathrm{E}[\xi_1^2(\xi_1^2-1)^2]\tilde{a}_{l,l}^2\,\mathrm{E}\left(\sum_{i=1}^{l-1}\tilde{a}_{i,l}\xi_i + \sum_{i=l+1}^{n}\tilde{a}_{i,l}z_i\right)^2 \\
&+ 16\,\mathrm{E}(\xi_1^4)\,\mathrm{E}\left(\sum_{i=1}^{l-1}\tilde{a}_{i,l}\xi_i + \sum_{i=l+1}^{n}\tilde{a}_{i,l}z_i\right)^4 \\
=& \mathrm{E}(\xi_1^2-1)^4\tilde{a}_{l,l}^4 + 24\,\mathrm{E}[\xi_1^2(\xi_1^2-1)^2]\tilde{a}_{l,l}^2\left((\sum_{i=1}^{n}\tilde{a}_{i,l}^2) - \tilde{a}_{l,l}^2\right) \\
&+ 16\,\mathrm{E}(\xi_1^4)\left(\big(\mathrm{E}(\xi_1^4)-3\big)\sum_{i=1}^{l-1}\tilde{a}_{i,l}^4 + 3\left((\sum_{i=1}^{n}\tilde{a}_{i,l}^2) - \tilde{a}_{l,l}^2\right)^2\right).
\end{aligned}$$

To upper bound the above quantity, we use the facts $24\,\mathrm{E}[\xi_1^2(\xi_1^2-1)^2] \leq 2(16\,\mathrm{E}(\xi_1^2-1)^4+(9/4)\,\mathrm{E}(\xi_1^4))$, $\mathrm{E}(\xi_1^2-1)^4 \leq \mathrm{E}(\xi_1^8)$ and

$$\big(\mathrm{E}(\xi_1^4)-3\big)\sum_{i=1}^{l-1}\tilde{a}_{i,l}^4 \leq \big(\mathrm{E}(\xi_1^4)-1\big)\sum_{i=1}^{l-1}\tilde{a}_{i,l}^4 \leq \big(\mathrm{E}(\xi_1^4)-1\big)\left((\sum_{i=1}^{n}\tilde{a}_{i,l}^2) - \tilde{a}_{l,l}^2\right)^2.$$

Then we obtain the bound

$$\mathrm{E}(h_l^4) \leq \big(16\,\mathrm{E}(\xi_1^8) + 32\,\mathrm{E}(\xi_1^4)\big)\left(\sum_{i=1}^{n}\tilde{a}_{i,l}^2\right)^2. \tag{13}$$

Similarly, we have

$$\mathrm{E}(g_l^4) \leq \big(48\,\mathrm{E}(\xi_1^4) + 3\tau^4 + 96\big)\left(\sum_{i=1}^{n}\tilde{a}_{i,l}^2\right)^2. \tag{14}$$

Combining (12), (13) and (14) yields

$$\begin{aligned}
&\sum_{l=1}^{n}|\mathrm{E}\,f(S_l+h_l) - \mathrm{E}\,f(S_l+g_l)| \\
\leq& \frac{\big|\mathrm{E}(\xi_1^4)-1-\tau^2\big|}{2}\|f^{(2)}\|_\infty\sum_{l=1}^{n}\tilde{a}_{l,l}^2 \\
&+ \frac{\max\big(\big|\mathrm{E}(\xi_1^2-1)^3\big|, 12(\mathrm{E}(\xi_1^4)-1)\big)}{6}\|f^{(3)}\|_\infty\sum_{l=1}^{n}\left(|\tilde{a}_{l,l}|\sum_{i=1}^{n}\tilde{a}_{i,l}^2\right) \\
&+ \frac{16\,\mathrm{E}(\xi_1^8) + 80\,\mathrm{E}(\xi_1^4) + 3\tau^4 + 96}{24}\|f^{(4)}\|_\infty\sum_{l=1}^{n}\left(\sum_{i=1}^{n}\tilde{a}_{i,l}^2\right)^2.
\end{aligned}$$

This completes the proof.

□

Proof of Theorem 2. Throughout the proof, we use the similar notations as in Theorem 1 and define

$$S = \frac{(\sqrt{\phi}\boldsymbol{\epsilon})^\top \mathbf{A}\sqrt{\phi}\boldsymbol{\epsilon} - \operatorname{tr}(\mathbf{A})}{\sqrt{2\operatorname{tr}(\mathbf{A}^2) + (\phi^2 \operatorname{E}(\epsilon_1^4) - 3)\operatorname{tr}(\mathbf{A}^{\circ 2})}}$$

and

$$S_{\hat{\tau}}^* = \frac{\hat{\tau}\sum_{i=1}^n a_{i,i}\check{z}_i + 2\sum_{1\leq i<j\leq n} a_{i,j}z_i z_j}{\sqrt{2\operatorname{tr}(\mathbf{A}^2) + (\phi^2 \operatorname{E}(\epsilon_1^4) - 3)\operatorname{tr}(\mathbf{A}^{\circ 2})}},$$

where $z_1, \ldots, z_n, \check{z}_1, \ldots, \check{z}_n$ are iid $\mathcal{N}(0,1)$ random variables and are independent of $\hat{\tau}^2$.

By a standard subsequence argument, we only need to prove the theorem along a subsequence of $\{n\}$. Hence, without loss of generality, we assume $\hat{\tau}^2 \xrightarrow{a.s.} \phi^2 \operatorname{E}(\epsilon_1^4) - 1$. Write

$$\begin{aligned} S_{\hat{\tau}}^* =& \frac{\sqrt{\phi^2 \operatorname{E}(\epsilon_1^4) - 1}\sum_{i=1}^n a_{i,i}\check{z}_i + 2\sum_{1\leq i<j\leq n} a_{i,j}z_i z_j}{\sqrt{2\operatorname{tr}(\mathbf{A}^2) + (\phi^2 \operatorname{E}(\epsilon_1^4) - 3)\operatorname{tr}(\mathbf{A}^{\circ 2})}} \\ &+ \frac{(\hat{\tau} - \sqrt{\phi^2 \operatorname{E}(\epsilon_1^4) - 1})\sum_{i=1}^n a_{i,i}\check{z}_i}{\sqrt{2\operatorname{tr}(\mathbf{A}^2) + (\phi^2 \operatorname{E}(\epsilon_1^4) - 3)\operatorname{tr}(\mathbf{A}^{\circ 2})}} \\ =:& S_{\hat{\tau},1}^* + S_{\hat{\tau},2}^*. \end{aligned}$$

Note that $S_{\hat{\tau},1}^*$ is independent of $\hat{\tau}$. Since $\operatorname{E}(S_{\hat{\tau},1}^{*2}) = 1$, the distributions $\mathcal{L}(S_{\hat{\tau},1}^*)$ are tight as $n \to \infty$. Hence, without loss of generality, we assume $\mathcal{L}(S_{\hat{\tau},1}^*)$ weakly converges to a limit distribution with distribution function $F^\dagger(x)$. Let $S^\dagger$ be a random variable with distribution function $F^\dagger(x)$. By some algebra (See, e.g., (Chen et al., 2010, Proposition A.1.(iii))), it can be shown that $\operatorname{E}(S_{\hat{\tau},1}^{*4})$ is uniformly bounded. Then $\mathcal{L}(S_{\hat{\tau},1}^{*2})$ is uniformly integrable. Hence $\operatorname{E}(S^{\dagger 2}) = 1$ and $F^\dagger(x)$ can not concentrate on a single point. Consequently, $F^\dagger(x)$ is continuous and is strict increasing for $x \in \{x : 0 < F(x) < 1\}$; see Sevast'yanov (1961) as well the remark made by A. N. Kolmogorov on that paper.

The condition (7) implies that $\operatorname{E}[S_{\hat{\tau},2}^{*2}|\hat{\tau}] \to 0$ almost surely. Then almost surely, $\mathcal{L}(S_{\hat{\tau}}^*|\hat{\tau}) \rightsquigarrow \mathcal{L}(S^\dagger)$. Consequently, for every $f \in \mathscr{C}^4(\mathbb{R})$, we have $|\operatorname{E}[f(S_{\hat{\tau}}^*)|\hat{\tau}] - \operatorname{E} f(S^\dagger)| \to 0$ almost surely. On the other hand, Theorem 1 and the condition (7) imply $|\operatorname{E} f(S) - \operatorname{E}[f(S_{\hat{\tau}}^*)|\hat{\tau}]| \to 0$ almost surely. Thus, $|\operatorname{E} f(S) - \operatorname{E} f(S^\dagger)| \to 0$. That is, $\mathcal{L}(S) \rightsquigarrow \mathcal{L}(S^\dagger)$.

Note that

$$x^{(1)} = \frac{F^{-1}(1-\alpha; \mathbf{A}, \hat{\tau}) - \operatorname{tr}(\mathbf{X}_b^* \mathbf{X}_b^{*\top})^{-1}}{n-q}.$$

We need to deal with $F^{-1}(1-\alpha; \mathbf{A}, \hat{\tau})$. Since $\mathcal{L}(S_{\hat{\tau}}^*|\hat{\tau}) \rightsquigarrow \mathcal{L}(S^\dagger)$ almost surely, the fact

$$\Pr\left(S_{\hat{\tau}}^* > \frac{F^{-1}(1-\alpha; \mathbf{A}, \hat{\tau})}{\sqrt{2\operatorname{tr}(\mathbf{A}^2) + (\phi^2 \operatorname{E}(\epsilon_1^4) - 3)\operatorname{tr}(\mathbf{A}^{\circ 2})}} \middle| \hat{\tau} \right) = \alpha$$

implies that almost surely,

$$\frac{F^{-1}(1-\alpha; \mathbf{A}, \hat{\tau})}{\sqrt{2\operatorname{tr}(\mathbf{A}^2) + (\phi^2 \operatorname{E}(\epsilon_1^4) - 3)\operatorname{tr}(\mathbf{A}^{\circ 2})}} \to F^{\dagger -1}(1-\alpha). \tag{15}$$

We also need the fact that

$$(\sqrt{\phi}\boldsymbol{\epsilon})^\top \tilde{\mathbf{U}}_a \tilde{\mathbf{U}}_a^\top (\sqrt{\phi}\boldsymbol{\epsilon}) = (1 + o_p(1))(n - q), \tag{16}$$

which is a consequence of

$$\mathrm{E}\left((\sqrt{\phi}\boldsymbol{\epsilon})^\top \tilde{\mathbf{U}}_a \tilde{\mathbf{U}}_a^\top (\sqrt{\phi}\boldsymbol{\epsilon})\right) = n - q, \quad \mathrm{Var}\left((\sqrt{\phi}\boldsymbol{\epsilon})^\top \tilde{\mathbf{U}}_a \tilde{\mathbf{U}}_a^\top (\sqrt{\phi}\boldsymbol{\epsilon})\right) = O(n - q).$$

The fact $S \rightsquigarrow S^\dagger$, the equations (15), (16) and Slutsky's theorem lead to

$$\begin{aligned}
&\Pr\left(T > x^{(1)}\right) \\
=&\Pr\left(T > \frac{F^{-1}(1-\alpha; \mathbf{A}, \hat{\tau}) - \operatorname{tr}\left((\mathbf{X}_b^* \mathbf{X}_b^{*\top})^{-1}\right)}{n - q}\right) \\
=&\Pr\left((\sqrt{\phi}\boldsymbol{\epsilon})^\top \mathbf{A} (\sqrt{\phi}\boldsymbol{\epsilon}) > \frac{(\sqrt{\phi}\boldsymbol{\epsilon})^\top \tilde{\mathbf{U}}_a \tilde{\mathbf{U}}_a^\top (\sqrt{\phi}\boldsymbol{\epsilon})}{n - q} F^{-1}(1-\alpha; \mathbf{A}, \hat{\tau})\right) \\
=&\Pr\left(S > \frac{(\sqrt{\phi}\boldsymbol{\epsilon})^\top \tilde{\mathbf{U}}_a \tilde{\mathbf{U}}_a^\top (\sqrt{\phi}\boldsymbol{\epsilon})}{n - q} \frac{F^{-1}(1-\alpha; \mathbf{A}, \hat{\tau})}{\sqrt{2\operatorname{tr}(\mathbf{A}^2) + (\phi^2 \mathrm{E}(\epsilon_1^4) - 3)\operatorname{tr}(\mathbf{A}^{\circ 2})}}\right) \\
=&\Pr\left(S > (1 + o_P(1)) F^{-1}(1-\alpha)\right) \\
\to&\alpha.
\end{aligned}$$

This proves the theorem.

□

Proof of Proposition 2. From (Bai et al., 2018, Theorem 2.1), one can obtain the explicit forms of $\mathrm{Var}\left(\tilde{\boldsymbol{\epsilon}}^\top \left(\tilde{\mathbf{P}}_a\right) \tilde{\boldsymbol{\epsilon}}\right)$ and $\mathrm{Var}\left(\sum_{i=1}^n \tilde{\epsilon}_i^4\right)$ which involves the traces of certain matrices. Using (Horn and Johnson, 1991, Theorem 5.5.1), one can see that the eigenvalues of these matrices are all bounded. Hence it can be deduced that $\mathrm{Var}(\tilde{\boldsymbol{\epsilon}}^\top \tilde{\mathbf{P}}_a \tilde{\boldsymbol{\epsilon}}) = O(n)$ and $\mathrm{Var}\left(\sum_{i=1}^n \tilde{\epsilon}_i^4\right) = O(n)$. Thus,

$$\begin{aligned}
&\tilde{\boldsymbol{\epsilon}}^\top \tilde{\mathbf{P}}_a \tilde{\boldsymbol{\epsilon}} = (n - q)\sigma^2 + O_P(\sqrt{n}), \\
&\sum_{i=1}^n \tilde{\epsilon}_i^4 = 3\sigma^4 \operatorname{tr}(\tilde{\mathbf{P}}_a^{\circ 2}) + \left(\mathrm{E}(\epsilon_1^4) - 3\sigma^4\right) \operatorname{tr}\left(\tilde{\mathbf{P}}_a^{\circ 2}\right)^2 + O_P(\sqrt{n}).
\end{aligned}$$

It follows that

$$\hat{\tau}^2 = \sigma^{-4}\mathrm{E}(\epsilon_1^4) - 1 + O_P\left(\frac{\sqrt{n}}{\operatorname{tr}\left(\tilde{\mathbf{P}}_a^{\circ 2}\right)^2}\right).$$

Let $\delta_{i,j} = 1$ if $i = j$ and 0 if $i \neq j$. We have

$$
\begin{aligned}
n =& \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_{i,j}^4 \\
=& \sum_{i=1}^{n} \sum_{j=1}^{n} \left((\tilde{\mathbf{P}}_a)_{i,j} + (\mathbf{P}_a)_{i,j} \right)^4 \\
\leq& 8 \sum_{i=1}^{n} \sum_{j=1}^{n} \left((\tilde{\mathbf{P}}_a)_{i,j} \right)^4 + 8 \sum_{i=1}^{n} \sum_{j=1}^{n} (\mathbf{P}_a)_{i,j}^4 \\
\leq& 8 \sum_{i=1}^{n} \sum_{j=1}^{n} \left((\tilde{\mathbf{P}}_a)_{i,j} \right)^4 + 8 \sum_{i=1}^{n} \sum_{j=1}^{n} (\mathbf{P}_a)_{i,j}^2 \\
=& 8 \operatorname{tr} \left(\tilde{\mathbf{P}}_a^{\circ 2} \right)^2 + 8q.
\end{aligned}
$$

Then

$$
\frac{\sqrt{n}}{\operatorname{tr} \left(\tilde{\mathbf{P}}_a^{\circ 2} \right)^2} = O\left(\frac{1}{\sqrt{n}} \right).
$$

This completes the proof. □

Proof of Proposition 3. Without loss of generality, we assume $\mathbf{A}$ is a diagonal matrix and $|b_1| \geq \cdots \geq |b_n|$. By a standard subsequence argument, we only need to prove the result along a subsequence. Hence we can assume $\lim_{n\to\infty} \|b\|^2 / \operatorname{tr}(\mathbf{A}^2) = c \in [0, +\infty]$. If $c = 0$, Lyapunov central limit theorem implies that

$$
\frac{Z^\top \mathbf{A} Z + b^\top Z - \operatorname{tr}(\mathbf{A})}{\sqrt{2 \operatorname{tr}(\mathbf{A}^2) + \|\mathbf{b}\|^2}} = (1 + o_P(1)) \frac{Z^\top \mathbf{A} Z - \operatorname{tr}(\mathbf{A})}{\sqrt{2 \operatorname{tr}(\mathbf{A}^2)}} + o_P(1) \rightsquigarrow \mathcal{N}(0,1).
$$

If $c = +\infty$,

$$
\frac{Z^\top \mathbf{A} Z + b^\top Z - \operatorname{tr}(\mathbf{A})}{\sqrt{2 \operatorname{tr}(\mathbf{A}^2) + \|\mathbf{b}\|^2}} = (1 + o_P(1)) \frac{b^\top Z}{\|\mathbf{b}\|} + o_P(1) \rightsquigarrow \mathcal{N}(0,1).
$$

In what follows, we assume $c \in (0, +\infty)$. By Helly selection theorem, we can assume $\lim_{n\to\infty} |b_i| / \|\mathbf{b}\| = b_i^* \in [0,1]$, $i = 1, 2, \ldots$. From Fatou's lemma, we have $\sum_{i=1}^{\infty} (b_i^*)^2 \leq 1$. Consequently, $\lim_{i\to\infty} b_i^* = 0$.

Note that the condition $\operatorname{tr}(\mathbf{A}^4) / \operatorname{tr}^2(\mathbf{A}^2) \to 0$ is equivalent to $\lambda_1(\mathbf{A}^2) / \operatorname{tr}(\mathbf{A}^2) \to 0$. Then for every fixed integer $r > 0$,

$$
\frac{\sum_{i=1}^{r} a_{i,i}^2}{\sum_{i=1}^{n} a_{i,i}^2} \leq \frac{r \max_{1 \leq i \leq n} a_{i,i}^2}{\sum_{i=1}^{n} a_{i,i}^2} \to 0.
$$

Then there exists a sequence of positive integers $r(n) \to \infty$ such that $\left(\sum_{i=1}^{r(n)} a_{i,i}^2 \right) / \left(\sum_{i=1}^{n} a_{i,i}^2 \right) \to 0$ and $r(n)/n \to 0$. Write

$$
Z^\top \mathbf{A} Z + b^\top Z - \operatorname{tr}(\mathbf{A}) = \sum_{i=1}^{r(n)} a_{i,i}(z_i^2 - 1) + \sum_{i=1}^{r(n)} b_i z_i + \sum_{i=r(n)+1}^{n} \left(a_{i,i}(z_i^2 - 1) + b_i z_i \right),
$$

which is a sum of independent random variables. The first term is negligible since $\mathrm{Var}(\sum_{i=1}^{r(n)} a_{i,i}(z_i^2 - 1)) = o(\sum_{i=1}^{n} a_{i,i}^2)$. Now we deal with the third term. From Berry-Esseen inequality (See, e.g., (DasGupta, 2008, Theorem 11.2)), there exists an absolute constant $C^* > 0$, such that

$$
\begin{aligned}
&\sup_{x\in\mathbb{R}} \left| \Pr\left(\frac{\sum_{i=r(n)+1}^{n} \left(a_{i,i}(z_i^2-1) + b_i z_i\right)}{\sqrt{2\sum_{i=r(n)+1}^{n} a_{i,i}^2 + \sum_{i=r(n)+1}^{n} b_i^2}} \leq x \right) - \Phi(x) \right| \\
\leq & C^* \frac{\sum_{i=r(n)+1}^{n} \mathrm{E}\left|a_{i,i}(z_i^2-1) + b_i z_i\right|^3}{\left(2\sum_{i=r(n)+1}^{n} a_{i,i}^2 + \sum_{i=r(n)+1}^{n} b_i^2\right)^{3/2}}.
\end{aligned}
$$

By some simple algebra, there exist absolute constants $C_1^*, C_2^* > 0$ such that for sufficiently large n,

$$
\begin{aligned}
&\sup_{x\in\mathbb{R}} \left| \Pr\left(\frac{\sum_{i=r(n)+1}^{n} \left(a_{i,i}(z_i^2-1) + b_i z_i\right)}{\sqrt{2\sum_{i=r(n)+1}^{n} a_{i,i}^2 + \sum_{i=r(n)+1}^{n} b_i^2}} \leq x \right) - \Phi(x) \right| \\
\leq & C_1^* \frac{\max_{1\leq i\leq n} |a_{i,i}|}{\sqrt{\sum_{i=1}^{n} a_{i,i}^2}} + C_2^* \frac{|b_{r(n)+1}|}{\|\mathbf{b}\|}.
\end{aligned}
$$

Since the right hand side tends to 0, we have

$$
\frac{\sum_{i=r(n)+1}^{n} \left(a_{i,i}(z_i^2-1) + b_i z_i\right)}{\sqrt{2\sum_{i=r(n)+1}^{n} a_{i,i}^2 + \sum_{i=r(n)+1}^{n} b_i^2}} \rightsquigarrow \mathcal{N}(0,1).
$$

Note that $\sum_{i=r(n)+1}^{n} \left(a_{i,i}(z_i^2-1) + b_i z_i\right)$ is independent of $\sum_{i=1}^{r(n)} b_i z_i$ and $\sum_{i=1}^{r(n)} b_i z_i \sim \mathcal{N}(0, \sum_{i=1}^{r(n)} b_i^2)$. Thus,

$$
\frac{\sum_{i=1}^{r(n)} b_i z_i + \sum_{i=r(n)+1}^{n} \left(a_{i,i}(z_i^2-1) + b_i z_i\right)}{\sqrt{2\sum_{i=1}^{n} a_{i,i}^2 + \sum_{i=1}^{n} b_i^2}} \rightsquigarrow \mathcal{N}(0,1).
$$

This completes the proof.

□

Note that under the normality, $T_n - \mathrm{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})/(n-q)$ has zero mean.

Proof of Theorem 3. We note that

$$
\begin{aligned}
&\Pr\left(\frac{\mathbf{y}^{*\top}\left(\mathbf{X}_b^*\mathbf{X}_b^{*\top}\right)^k \mathbf{y}^*}{\mathbf{y}^{*\top}\mathbf{y}^*} \leq \mathrm{E}(\gamma_I^k) + \sqrt{\frac{2\,\mathrm{Var}\left(\gamma_I^k\right)}{n-q}} x \right) \\
=&\Pr\left(\mathbf{y}^{*\top}\left(\mathbf{X}_b^*\mathbf{X}_b^{*\top}\right)^k \mathbf{y}^* \leq \left(\mathrm{E}(\gamma_I^k) + \sqrt{\frac{2\,\mathrm{Var}\left(\gamma_I^k\right)}{n-q}} x \right) \mathbf{y}^{*\top}\mathbf{y}^* \right) \qquad (17) \\
=&\Pr\left(\mathbf{y}^{*\top}\mathbf{B}\mathbf{y}^* \leq 0 \right),
\end{aligned}
$$

where

$$\mathbf{B} = \left(\mathbf{X}_b^* \mathbf{X}_b^{*\top}\right)^k - \left(\mathrm{E}(\gamma_I^k) + \sqrt{\frac{2\,\mathrm{Var}\left(\gamma_I^k\right)}{n-q}}x\right)\mathbf{I}_{n-q}.$$

Since $\mathbf{y}^{*\top}\mathbf{B}\mathbf{y}^* = \boldsymbol{\epsilon}^\top \tilde{\mathbf{U}}_a \mathbf{B} \tilde{\mathbf{U}}_a^\top \boldsymbol{\epsilon} + 2\boldsymbol{\epsilon}^\top \tilde{\mathbf{U}}_a \mathbf{B}\mathbf{X}_b^* \boldsymbol{\beta}_b + \boldsymbol{\beta}_b^\top \mathbf{X}_b^{*\top}\mathbf{B}\mathbf{X}_b^*\boldsymbol{\beta}_b$, we have

$$\begin{aligned}
&\Pr\left(\mathbf{y}^{*\top}\mathbf{B}\mathbf{y}^* \leq 0\right) \\
=&\Pr\left(\frac{\boldsymbol{\epsilon}^\top \tilde{\mathbf{U}}_a \mathbf{B} \tilde{\mathbf{U}}_a^\top \boldsymbol{\epsilon} + 2\boldsymbol{\epsilon}^\top \tilde{\mathbf{U}}_a \mathbf{B}\mathbf{X}_b^* \boldsymbol{\beta}_b - \phi^{-1}\,\mathrm{tr}(\mathbf{B})}{\sqrt{2\phi^{-2}\,\mathrm{tr}(\mathbf{B}^2) + 4\phi^{-1}\boldsymbol{\beta}_b^\top \mathbf{X}_b^{*\top}\mathbf{B}^2\mathbf{X}_b^*\boldsymbol{\beta}_b}}\right. \\
&\qquad\left.\leq \frac{-\boldsymbol{\beta}_b^\top \mathbf{X}_b^{*\top}\mathbf{B}\mathbf{X}_b^*\boldsymbol{\beta}_b - \phi^{-1}\,\mathrm{tr}(\mathbf{B})}{\sqrt{2\phi^{-2}\,\mathrm{tr}(\mathbf{B}^2) + 4\phi^{-1}\boldsymbol{\beta}_b^\top \mathbf{X}_b^{*\top}\mathbf{B}^2\mathbf{X}_b^*\boldsymbol{\beta}_b}}\right).
\end{aligned}$$

To apply proposition 3, we need to verify the condition $\lambda_1\left(\mathbf{B}^2\right)/\,\mathrm{tr}\left(\mathbf{B}^2\right) \to 0$. It is straightforward to show that $\mathrm{tr}(\mathbf{B}^2) = (n-q+2x^2)\,\mathrm{Var}(\gamma_I^k)$. On the other hand,

$$\begin{aligned}
\lambda_1\left(\mathbf{B}^2\right) =& \max_{1\leq i\leq n-q}\left(\gamma_i^k - \mathrm{E}(\gamma_I^k) - \sqrt{\frac{2\,\mathrm{Var}\left(\gamma_I^k\right)}{n-q}}x\right)^2 \\
\leq& 2\max_{1\leq i\leq n-q}\left(\gamma_i^k - \mathrm{E}(\gamma_I^k)\right)^2 + 4\frac{\mathrm{Var}\left(\gamma_I^k\right)}{n-q}x^2.
\end{aligned}$$

Thus,

$$\begin{aligned}
\frac{\lambda_1(\mathbf{B}^2)}{\mathrm{tr}(\mathbf{B}^2)} \leq& 2\frac{\max_{1\leq i\leq n-q}\left(\gamma_i^k - \mathrm{E}(\gamma_I^k)\right)^2}{(n-q+2x^2)\,\mathrm{Var}(\gamma_I^k)} + 4\frac{x^2}{(n-q)(n-q+x^2)} \\
\leq& 2\frac{\max_{1\leq i\leq n-q}\left(\gamma_i^k - \mathrm{E}(\gamma_I^k)\right)^2}{(n-q)\,\mathrm{Var}(\gamma_I^k)} + \frac{4}{(n-q)},
\end{aligned}$$

which tends to 0 by the condition (8). Hence Proposition 3 implies that

$$\Pr\left(\mathbf{y}^{*\top}\mathbf{B}\mathbf{y}^* \leq 0\right) = \Phi\left(\frac{-\boldsymbol{\beta}_b^\top \mathbf{X}_b^{*\top}\mathbf{B}\mathbf{X}_b^*\boldsymbol{\beta}_b - \phi^{-1}\,\mathrm{tr}(\mathbf{B})}{\sqrt{2\phi^{-2}\,\mathrm{tr}(\mathbf{B}^2) + 4\phi^{-1}\boldsymbol{\beta}_b^\top \mathbf{X}_b^{*\top}\mathbf{B}^2\mathbf{X}_b^*\boldsymbol{\beta}_b}}\right) + o(1). \tag{18}$$

Then the conclusion follows from (17), (18) and the following facts

$$\begin{aligned}
\mathrm{tr}(\mathbf{B}) =& -(n-q)\sqrt{\frac{2\,\mathrm{Var}\left(\gamma_I^k\right)}{n-q}}x, \\
\mathrm{tr}(\mathbf{B}^2) =& (1+o(1))(n-q)\,\mathrm{Var}(\gamma_I^k), \\
\boldsymbol{\beta}_b^\top \mathbf{X}_b^{*\top}\mathbf{B}\mathbf{X}_b^*\boldsymbol{\beta}_b =& (n-q)\left(\mathrm{Cov}\left(\gamma_I^k, \gamma_I w_I^2\right) - \mathrm{E}(\gamma_I w_I^2)\sqrt{\frac{2\,\mathrm{Var}\left(\gamma_I^k\right)}{n-q}}x\right), \\
\boldsymbol{\beta}_b^\top \mathbf{X}_b^{*\top}\mathbf{B}^2\mathbf{X}_b^*\boldsymbol{\beta}_b =& (n-q)\,\mathrm{E}\left[\left(\gamma_I^k - \mathrm{E}(\gamma_I^k) - \sqrt{\frac{2\,\mathrm{Var}(\gamma_I^k)}{n-q}}x\right)^2\gamma_I w_I^2\right].
\end{aligned}$$

□

References

Bai, Z., Pan, G., and Yin, Y. (2018). A central limit theorem for sums of functions of residuals in a high-dimensional regression model with an application to variance homoscedasticity test. *TEST*, 27(4):896–920.

Bentkus, V. and Götze, F. (1996). Optimal rates of convergence in the CLT for quadratic forms. *The Annals of Probability*, 24(1):466–490.

Bühlmann, P. (2013). Statistical significance in high-dimensional linear models. *Bernoulli*, 19(4):1212–1242.

Casella, G. and Moreno, E. (2006). Objective bayesian variable selection. *Journal of the American Statistical Association*, 101(473):157–167.

Chatterjee, S. (2008). A new method of normal approximation. *The Annals of Probability*, 36(4):1584–1610.

Chen, S. X., Zhang, L.-X., and Zhong, P.-S. (2010). Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association*, 105(490):810–819.

Cohn, D. L. (2013). *Measure Theory*. Birkhäuser, New York, 2nd edition.

DasGupta, A. (2008). *Asymptotic Theory of Statistics and Probability*. Springer, New York, 1st edition.

Dezeure, R., Bühlmann, P., and Zhang, C.-H. (2017). High-dimensional simultaneous inference with the bootstrap. *TEST*, 26(4):685–719.

Dicker, L. H. and Erdogdu, M. A. (2017). Flexible results for quadratic forms with applications to variance components estimation. *The Annals of Statistics*, 45(1):386–414.

Fan, J., Yuan, L., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(4):603–680.

Feng, L., Zou, C., Wang, Z., and Chen, B. (2013). Rank-based score tests for high-dimensional regression coefficients. *Electronic Journal of Statistics*, 7(0):2131–2149.

Goddard, S. D. and Johnson, V. E. (2016). Restricted most powerful bayesian tests for linear models. *Scandinavian Journal of Statistics*, 43(4):1162–1177.

Goeman, J. J., van de Geer, S. A., and van Houwelingen, H. C. (2006). Testing against a high dimensional alternative. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):477–493.

Goeman, J. J., van Houwelingen, H. C., and Finos, L. (2011). Testing against a high-dimensional alternative in the generalized linear model: asymptotic type i error control. *Biometrika*, 98(2):381–390.

Götze, F. and Tikhomirov, A. (2002). Asymptotic distribution of quadratic forms and applications. *Journal of Theoretical Probability*, 15(2):423–475.

Horn, R. A. and Johnson, C. R. (1991). *Topics in Matrix Analysis*. Cambridge University Press, New York, 1st edition.

Janson, L., Barber, R. F., and Candès, E. (2016). EigenPrism: inference for high dimensional signal-to-noise ratios. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(4):1037–1065.

Javier Girón, F., Lina Martínez, M., Moreno, E., and Torres, F. (2006). Objective testing procedures in linear models: Calibration of the p-values. *Scandinavian Journal of Statistics*, 33(4):765–784.

Jiang, J. (1996). Reml estimation: asymptotic behavior and related topics. *Annals of Statistics*, 24(1):255–286.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.

Lan, W., Ding, Y., Fang, Z., and Fang, K. (2016a). Testing covariates in high dimension linear regression with latent factors. *Journal of Multivariate Analysis*, 144:25–37.

Lan, W., Wang, H., and Tsai, C. L. (2014). Testing covariates in high-dimensional regression. *Annals of the Institute of Statistical Mathematics*, 66(2):279–301.

Lan, W., Zhong, P.-S., Li, R., Wang, H., and Tsai, C.-L. (2016b). Testing a single regression coefficient in high dimensional linear models. *Journal of Econometrics*, 195(1):154–168.

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. Springer, New York, 3rd edition.

Lei, L., Bickel, P. J., and Karoui, N. E. (2018). Asymptotics for high dimensional regression m-estimates: fixed design results. *Probability Theory and Related Fields*, 172(3-4):983–1079.

Liang, F., Paulo, R., Molina, G., Clyde, M. A., and Berger, J. O. (2008). Mixtures of g priors for bayesian variable selection. *Journal of the American Statistical Association*, 103(481):410–423.

Pollard, D. (1984). *Convergence of stochastic processes*. Springer, New York, 1st edition.

Sevast'yanov, B. A. (1961). A class of limit distributions for quadratic forms of stochastic random variables. *Theory of Probability & Its Applications*, 6(3):337–340.

Wang, S. and Cui, H. (2015). A new test for part of high dimensional regression coefficients. *Journal of Multivariate Analysis*, 137:187–203.

Xu, K. (2016). A new nonparametric test for high-dimensional regression coefficients. *Journal of Statistical Computation and Simulation*, 87(5):855–869.

Zhang, C.-H. and Zhang, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):217–242.

Zhong, P.-S. and Chen, S. X. (2011). Tests for high-dimensional regression coefficients with factorial designs. *Journal of the American Statistical Association*, 106(493):260–274.

Zhou, Q. and Guan, Y. (2018). On the null distribution of bayes factors in linear regression. *Journal of the American Statistical Association*, 113(523):1362–1371.